AGE-RELATED VARIATION OF PLOSIVE VOICE ONSET TIME IN STANDARD CHINESE

LI Qian

Abstract: It is conventionally believed that, once one has acquired his native phonemes, the acoustic realizations are relatively stable since the later stage of childhood, and of high intra-speaker consistency throughout adulthood. With evidence from plosives produced in connected speech in Standard Chinese, the present study shows that even middle-aged speakers show gradually increased voice onset time (VOT) in the production of plosives in general. Therefore, at least for phonemes such as plosives, the acoustic realizations are further developed and in an ongoing, dynamic process of change even in adulthood. Furthermore, by considering both raw VOT and VOT ratios, the present study also shows that the increase in VOT cannot be completely due to decrease in speech rate as speakers age, but could be related to speakers' adjustments to the physiological changes due to aging in speech production.

Keywords: Aging, Plosive, VOT, Standard Chinese

1. INTRODUCTION

It is known that speech sounds vary extensively in our daily communication which can be attributed to various sources. To understand these variations from the synchronic point of view, considerable research efforts have been devoted to contextual speech variations in both segmental (see [14] for a review) and suprasegmental aspects (see [11] for a review). From the diachronic perspective, the variations of speech sounds have been mainly accounted for by sociolinguistic factors such as generational differences (see [36] for a review). What is somewhat overlooked, however, is the fact that a speaker cannot even produce the same sound in the same way at different time points across the lifespan once he acquires his native language in childhood, notably due to aging. The acoustic realizations of native phonemes have been conventionally considered to be relatively stable from the later stage of childhood, and therefore of high intra-speaker consistency throughout adulthood, when the plasticity of articulators and brain is believed to be declining [31].

Such an understanding, however, has been challenged by evidence from studies on speech production by healthy aging adults. A number of studies have shown that, the phonetic realizations of one's native speech sounds are not unchanged in different stages of adulthood. As compared to the younger adults, the elderly show significant differences in multiple aspects of speech (see e.g., [37] for a review). Among them, age-related changes in the voice onset time (VOT) of plosives have been well noticed.

VOT is known as a robust acoustic cue for plosive consonants, which is defined as the interval between the release of articulatory occlusion and the onset of voicing for the following voiced segment, although the specific range of VOT for each plosive varies from one language to another [23]. It is resulted from a series of complex coordination of the lungs, the vocal folds as well as the supraglottal articulators in specific ways in different languages (e.g., [12]). The ability and the actual performance of precise controlling of all the above-mentioned organs are known to be developed gradually through childhood (e.g., [15, 19, 40]). Previous studies also show that, the VOT of plosives produced by aged adults exhibit different VOT patterns from those produced by younger adults (but see [24, 25, 35] for no significant difference between the two groups), possibly due to less efficient motor control (e.g., [20]) and degraded cognitive ability to process temporal information (e.g., [38]). However, no consensus has been reached so far on how exactly age affects the values of VOT in different studies.

On the one hand, various studies have shown that aged adults exhibit significantly shorter VOT than young adults (e.g., [4, 5, 17, 18, 20, 37]). For example, [20] compared the VOT values in three pairs of plosives in English (/b-p/, /d-t/, /g-k/) produced by one group of young male speakers (aged 25-39) with two aged male speaker groups (aged 60-69 and 70-79). Results show significant reduction on VOT in the production of two older speaker groups as compared to the young control group, especially for bilabial plosive (i.e., /p/ to be specific) and voiceless plosive (i.e., /t/ specifically). [37] showed similar trend for their aged subjects (aged 60-89) as compared to the younger subjects (aged 20-35), as their older adults had a significantly shorter mean VOT for /g/ compared to the younger subjects. Moreover, gender also seems to interact with the age effect in addition to a general gender main effect, specifically for the voiceless stop /t/, where the old male group had a significantly shorter mean VOT compared with the old female, young male, and young female groups. [18] also showed negative correlations between age and VOT of plosives in Czech, which, however, was only observed in female speakers.

On the other hand, there are also studies which show clear age effect towards the opposite direction, with lengthened VOT as speakers' age increase (e.g., [8, 10, 13]). For example, with evidence from a longitudinal study, [13] showed that subjects in their 60s produced significantly prolonged VOT of /p/ and /k/ in Dutch than those produced in their 30s; and the mean VOT values at two different age stages relatively corresponded to the mean VOT produced by subjects in similar age groups in another cross-sectional experiment in their paper (at least for /p/), although the differences among different age groups in the cross-sectional study failed to reach significance in most cases. In an investigation of plosives in Hungarian, [8] also showed that older speakers (aged 70-90) produced longer VOT for the bilabials and alveolars than the younger subjects (aged 21-32), although for the velars, the aged group exhibited shorter VOT. Furthermore, in studies that did not find significant VOT changes in the older subjects in general, longer VOTs in the older groups were nevertheless found in certain phonetic contexts, e.g., for velars followed by the vowel /a/ in [24].

It is noticeable that, for studies where older speakers were found to have produced shorter VOTs than younger speakers, the results were attributed to "slower motoric movements, weaker muscles, and other anatomical and physiological changes in the larynx and pharynx" [17], "reduced motoric efficiency" [20], or "diminishing of neuromuscular coordination and neural timing" [35] in older subjects. These, however, do not necessarily predict a shorter VOT value. Given the findings that multiple changes have taken place as people age, such as longer motor reaction time [9], slower rate of neural impulse conduction [1, 6], and slower speech diadochokinetic rate [27], longer VOT values of plosives are highly expected as the subject age increases. Note that in previous studies, limited reading materials have been used, e.g., a very small number of words produced in isolation; we could hardly exclude the possibility that results in those studies were affected by the characteristics of individual material items. Therefore, one of the main purposes of the current study is to revisit the effect of age on plosive VOT with inclusion of much larger number of material items.

Furthermore, regardless of the general direction of change, one issue which is not yet clear in the literature is how such age-related speech variation is developed over time throughout adulthood. It can be seen that in all studies that have been mentioned above, conclusions were generally drawn from comparisons between different non-continuous age groups, i.e., younger adulthood (typically 20s-30s) vs. older adulthood (typically above 60s), including in the only longitudinal study (i.e., [13]). It is therefore unknown what has happened in between, i.e., during the middle-aged stage, which has greatly limited our understanding of the development of speech production and perception across the lifespan and further modeling of speech variation in general.

It is worth noting that, with evidence from a non-speech visuomotor tracking task, [2] showed that, with the increase of age, the changes in the control of the lower lip, jaw, and larynx tended to decline at the earliest from the age of 30 in a continuous and gradual matter and well extend into the age of 80. The coordination of these has been assumed as being closely related to the VOT values in plosive productions [12]. Furthermore, in a recent study on vowel production, [16] also showed that the vowel formants also shifted in a gradual manner towards the periphery of the vowel space with the increase of speaker age from the age of 21 to 49, during which vowel production has conventionally been expected to be “stable”. These have pointed to the possibility that a similar gradient and continuous change in VOT might also be observed in the middle-aged period. The second goal of the current study thus aims to test this hypothesis. If any age-related change in VOT indeed happen within the middle-aged stage, we would expect to observe gradual VOT changes at least since the age of 30 (following [2]) and throughout the middle-age adulthood; otherwise, VOT with no significant age-related change would be observed during this period.

To recapitulate, there are two research questions in the current study concerning the VOT of plosives: 1) whether does VOT change as a function of age during the middle-aged adulthood? 2) what is the direction of the age-related change if any?

To answer these two questions, six plosives in Standard Chinese were examined in the current study. There are three pairs of voiceless plosives in Standard Chinese differentiated in the places of articulation, which occur unexceptionally at the onset of the syllables [21]: bilabial (i.e., /p/ as in *ba*^{1} /pa/ “eight” vs. /p^h/ as in *pa*^{1} /p^{h}a/ “to lie prone”), alveolar (i.e., /t/ as in *da*^{1} /ta/ “to build” vs. /t^h/ as in *ta*^{1} /t^{h}a/ “he/she”), and velar (i.e., /k/ as in *ge*^{1} /kɤ/ “song” vs. /k^h/ as in *ke*^{1} /k^{h}ɤ/ “subject”). Most previous studies on the aging effect on plosive VOT have mainly used stimuli in languages such as English, plosives in which are claimed to be differentiated in the voicing dimensions but not necessarily always exhibit negative-positive contrast in the VOT values (see [23]). These might potentially lead to difficulties in the interpretation of the data. The use of voiceless consonants with exclusively positive VOT thus better solves such problem.

In addition, as VOT might be influenced by speech rate, both raw VOT and VOT ratios were examined in the current study following [18]. Most of the previous studies did not take speech rate into serious consideration in the development of VOT at all. The introduction of VOT ratios therefore not only eliminates the individual differences in speech rate that further gives rise to considerable individual differences in VOT, but also excludes the possibility that any potential increase in VOT with age is correlated to decreased speech rate as the speakers get older.

2. METHODS

2.1 Participants

Thirty-three middle-aged adult speakers in total participated in the current study ranging in age from 32 to 50 (17 females: 39.4±5.0 yrs; 16 males: 40.4±6.4 yrs). The participants were native Northern Mandarin speakers. Based on ratings by one professional judge, all participants reached comparable proficiency levels of Standard Chinese (relatively as B level, i.e., good pronunciation with very slight dialectal accents). The education levels of the participants were also controlled to be senior high school or equivalent. No participants reported hearing impairments or any history for speech disorders.

2.2 Materials

All reading materials were sentences selected from a Mandarin speech corpus RASC863. The sentences were designed based on scripts of newspapers, interviews, and regular conversations (see [39] for further details). Each participant was spontaneously allocated with one set of these sentences (95 sentences in each set), so that the reading materials for each participant covered all phonemic and lexical tone combinations in Standard Chinese.

Among sentences produced by the participants, five sentences produced by two participants were excluded due to obvious reading mistakes. Similar distributions of plosives were produced by each participant, i.e., 101±13 items for /p/; 35±9 items for /p^h/; 122±17 items for /t/; 71±13 items for /t^h/; 90±12 items for /k/; 46±10 items for /k^h/.

2.3 Recordings

All participants were requested to read aloud the sentences presented on the computer screen at their normal speech rates. All found the task straightforward and followed the same procedure. Recordings were conducted using an M-Audio® sound card through a Sennheiser® headset microphone with a custom-written recording software in a quiet room. All data were collected with 16kHz sampling rate and 16-bit rate in mono channel.

2.4 Data analysis

The acoustic data were manually segmented in Praat [7]. VOT was defined as the interval between the consonant release burst and the vowel onset based on both spectrograms and oscillographic sound pressure signals displayed. If multiple bursts occurred, the initial burst was used. A custom-written script was used for VOT extraction. VOT ratios were further calculated as dividing the raw VOT value by the duration of whole syllable the plosive occurs in.

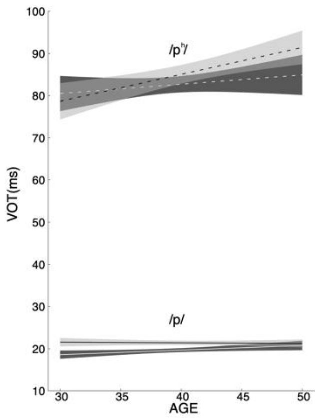
The linear mixed-effects modelling was used for statistics with the package lme4 [3] in R [28]. Models for raw VOT and VOT ratio were established separately. As multiple previous studies have shown a significant effect of phonetic contexts (i.e., vowels following the plosives) on VOT (e.g., [20, 24, 37]), RHYME was considered as a random effect in both models. For each dataset, the base model was constructed with only AGE in the fixed-factor structure and the random intercept of RHYME in the random-factor structure. Centering of the AGE factor was done throughout the current study for better interpretations of the results of the linear mixed-effects models. SUBJECT, however, was not considered in the random-factor structure as usual due to strong collinearity with AGE in both cases, considering which in the model might lead to unnecessary Type II errors. Other fixed factors as well as their interactions were added into the model in a stepwise fashion. Model fits were tested at each step via model comparisons based on the change in the log-likelihood ratios. Factors that significantly improved the model fits were kept in the fixed-factor structure of the model. Parameter-specific p-values were estimated using the normal approximation (i.e., treating the t-values as a z-value).

3. RESULTS

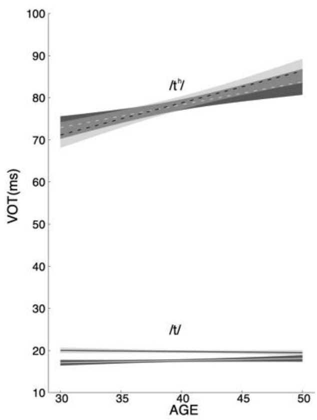
3.1 Raw VOT

Figure 1 illustrates the change of raw VOT of different plosives as a function of age using simple linear regression with 95% confidence regions. Figure 1*a* shows the VOT changes for bilabial plosives /p/ (see solid lines in Figure 1*a*) and /p^h/ (see dashed lines in Figure 1*a*), Figure 1*b* for alveolar plosives /t/ (see solid lines in Figure 1*b*) and /t^h/ (see dashed lines in Figure 1*b*), and Figure 1*c* for velar plosives /k/ (see solid lines in Figure 1*c*) and /k^h/ (see dashed lines in Figure 1*c*). Shades were used to illustrate possible VOT change trajectories with 95% confidence regions for different genders of speakers, i.e., dark gray for female and light gray for male.

a. BILABIAL



b. ALVEOLAR



c. VELAR

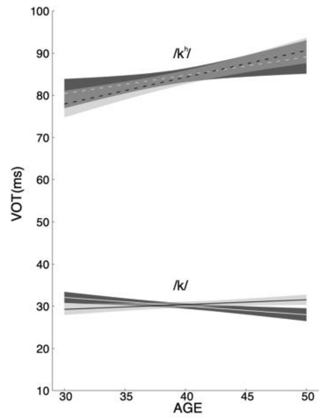


Figure 1: Simple linear regression fit with 95% confidence regions of the raw VOT (in milliseconds) of plosives as a function of age in different places of articulation, *a* for bilabial, *b* for alveolar, *c* for velar. Lines stand for the regression fits, solid lines for the unaspirated plosives and dashed lines for the aspirated ones. Shaded areas stand for the 95% confidence intervals around the regression fits, dark gray for female and light gray for male.

As can been seen from Figure 1 that, the change of raw VOT values with age show different patterns dependent on place of articulation, aspiration, the gender of the speakers. To be specific, all aspirated plosives, i.e., /p^h/, /t^h/, /k^h/ (see dashed lines in Figures 1*a*, 1*b*, 1*c*, respectively), show clear positive correlations between the raw VOT values and the speaker age, so that the VOT values increases as the speakers get older in general. This can be observed regardless of speaker genders. However, similar exclusively rising trends cannot be observed for unaspirated plosives, i.e., /p/, /t/, /k/ (see solid lines in Figures 1*a*, 1*b*, 1*c*, respectively). Compared to the aspirated plosives, the unaspirated ones show changes in raw VOT values in a much slighter manner, if any. Furthermore, /p/ and /t/ show similar patterns of VOT changes, where male speakers show declining VOT while the VOT for female speakers show increasing trends; /k/, however, shows exactly opposite direction of changes from those for /p/ and /t/, where male speakers show increasing VOT while female speakers show decreasing VOT.

Results of the linear mixed-effects modeling over the full dataset of the raw VOT values showed that ASPIRATION (two levels: aspirated vs. unaspirated), PLACE (three levels: bilabial, alveolar, and velar), GENDER (two levels: female vs. male), as well as their interactions (i.e., two-way interactions: AGE:ASPIRATION, AGE:GENDER, ASPIRATION:PLACE, ASPIRATION:GENDER, PLACE:GENDER; three-way interaction: ASPIRATION:PLACE:GENDER) significantly improved the model fits. These indicate that the raw values of VOT are predicted not only by the overall effects of the age and gender of the speakers, as well as the aspiration status and the place of articulation of plosives, but also by the complex interactions of those factors. Given this situation, the full dataset was then separated according to different plosives to better interpret the effect estimate of AGE on the raw VOT values. For each sub-dataset, the linear mixed-effects model was built in a similar manner as that for the full dataset. Table 1 lists the final full model for each sub-dataset and estimates of the fixed effects in each sub-model, in which *a* for /p/, *b* for /p^h/, *c* for /t/, *d* for /t^h/, *e* for /k/, and *f* for /k^h/. The estimate values of the fixed factor AGE in Table 1 stand for the slope of VOT change as a function of age.

Table 1: The final model and parametric estimates of raw VOT for different plosives, *a* for /p/, *b* for /p^h/, *c* for /t/, *d* for /t^h/, *e* for /k/, and *f* for /k^h/.

a. Final model for /p/:
AGE+GENDER+AGE:GENDER+(1|RHYME)
(Baseline for GENDER: FEMALE)

	β	*SE*	*t*	*p*
(Intercept)	19.19	0.62	30.93	<0.001
AGE	0.11	0.05	2.35	<0.05
GENDER	1.64	0.34	4.87	<0.001
AGE: GENDER	-0.13	0.06	-2.08	<0.05

b. Final model for /p^h/:
AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	82.87	2.34	35.40	<0.001
AGE	0.53	0.12	4.34	<0.001

c. Final model for /t/:
AGE+GENDER+AGE:GENDER+(1|RHYME)
(Baseline for GENDER: FEMALE)

	β	*SE*	*t*	*p*
(Intercept)	17.58	0.36	49.50	<0.001
AGE	0.06	0.03	1.70	n.s.
GENDER	2.26	0.23	9.77	<0.001
AGE: GENDER	-0.09	0.04	-2.20	<0.05

d. Final model for /t^h/:
AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	79.30	1.74	45.69	<0.001
AGE	0.73	0.08	9.20	<0.001

e. Final model for /k/:
AGE+AGE:GENDER+(1|RHYME)
(Baseline for GENDER: FEMALE)

	β	*SE*	*t*	*p*
(Intercept)	28.67	0.85	33.80	<0.001
AGE	-0.18	0.06	-2.79	<0.01
AGE: GENDER	0.29	0.08	3.46	<0.001

f. Final model for /k^h/:
AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	81.70	1.54	53.18	<0.001
AGE	0.60	0.10	5.92	<0.001

What can be seen from Table 1 is that, AGE shows significant main effects in all aspirated plosives (see Table 1*b*, 1*d*, 1*f*) where AGE is the only fixed-factor in the models; for the unaspirated plosives where models contain interactions of AGE and GENDER in the fixed-factor structure, significant simple effects of AGE can be observed except for /t/ (see Table 1*c*). In addition, significant simple intercept effects of GENDER can only be observed for /p/ (see Table 1*a*) and /t/ (see Table 1*c*), but not for /k/. To zoom into the effect of AGE, the estimated slope of AGE for each plosive were then computed based on the estimates in Table 1 and presented in Table 2. Positive values in the table indicate increasing VOT values as a function of age, while negative values indicate declining VOT values. * stands for estimated values significantly different from 0 (i.e., no correlation between VOT and age), indicating that the change in VOT is significant.

Table 2: The estimated slope of AGE for each plosive. * stands for estimated values significantly different from 0.

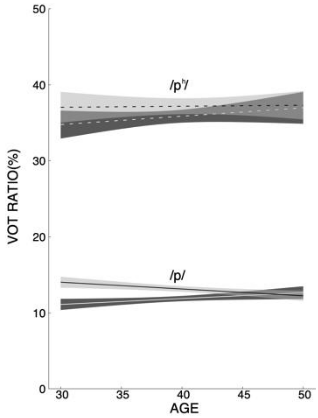
	Female	Male
/p/	0.11 *	-0.02 *
/pʰ/	0.53 *	
/t/	0.06	-0.03 *
/tʰ/	0.73 *	
/k/	-0.18 *	0.11 *
/kʰ/	0.60 *	

As can be seen from Table 2 that by and large, the estimated slopes of AGE for different plosives confirm what has been observed in Figure 1. To be specific, age shows exclusively positive slopes across genders in all aspirated plosives (see estimated slopes of AGE for /pʰ/, /tʰ/, /kʰ/ in Table 2); in unaspirated plosives, /p/ and /t/ show positive slopes of AGE for female speakers but negative slopes for male speakers although the slope of /t/ for female speakers is not significant (see estimated slopes of AGE for /p/ and /t/ in Table 2), whereas /k/ exhibits an opposite pattern where female speakers show a negative slope while male speakers show a positive slope (see estimated slopes of AGE for /k/ in Table 2). The magnitudes of estimated slopes are generally bigger for all aspirated plosives (>0.5), indicating relatively rapid changes of VOT as a function of age, while the unaspirated plosives show much smaller magnitudes of slopes regardless of the direction of change (<|0.5|).

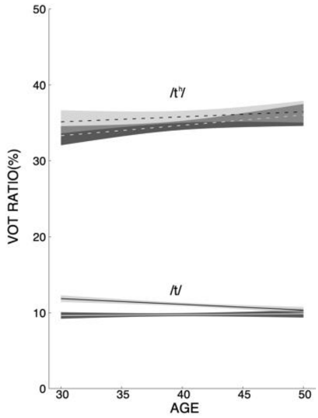
3.2 VOT Ratio

To tease apart whether changes in raw VOT values are co-varied with speech rate, we further examined the VOT ratios for each plosive and established comparable linear mixed-effects models as those for the raw VOT dataset. Figure 2 illustrates the change of VOT ratio of different plosives as a function of age using simple linear regression with 95% confidence regions. Figure 2*a* shows the VOT ratio changes for bilabial plosives /p/ (see solid lines in Figure 2*a*) and /pʰ/ (see dashed lines in Figure 2a), Figure 2*b* for alveolar plosives /t/ (see solid lines in Figure 2*b*) and /tʰ/ (see dashed lines in Figure 2*b*), and Figure 1*c* for velar plosives /k/ (see solid lines in Figure 2*c*) and /kʰ/ (see dashed lines in Figure 2*c*). Shades were used to illustrate possible VOT change trajectories with 95% confidence regions for different genders of speakers, i.e., dark gray for female and light gray for male.

a. BILABIAL



b. ALVEOLAR



c. VELAR

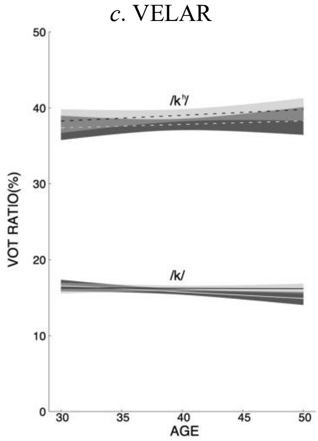


Figure 2: Simple linear regression fit with 95% confidence regions of the VOT ratios (in percentage) of plosives as a function of age in different places of articulation, *a* for bilabial, *b* for alveolar, *c* for velar. Lines stand for the regression fits, solid lines for the unaspirated plosives and dashed lines for the aspirated ones. Colored areas stand for the 95% confidence intervals around the regression fits, dark gray for female and light gray for male.

Compared to Figure 1, the changes in the VOT ratios in Figure 2 show similar trends as those of raw VOT values but with smaller magnitudes. As can be seen from Figure 2, all aspirated plosives, i.e., /p^h/, /t^h/, /k^h/ (see dashed lines in Figures 2*a*, 2*b*, 2*c*, respectively), show positive correlations between VOT ratio and speaker age across genders, but in much smaller magnitudes than those in Figure 1 (see dashed lines in Figure 1*a*, 1*b*, 1*c*, respectively).

For unaspirated plosives, i.e., /p/, /t/, /k/ (see solid lines in Figures 2*a*, 2*b*, 2*c*, respectively), the directions of change in VOT ratio are dependent on genders and places of articulation in similar ways as comparable ones in raw VOT values as in Figure 1 (see solid lines in Figures 1*a*, 1*b*, 1*c*, respectively), where male speakers show declining VOT ratios while female speakers show increasing trends for /p/ and /t/; /k/ shows slightly increasing VOT ratios for male speakers while female speakers show decreasing VOT ratios, but the differences between genders are less observable than those in the raw VOT values in Figure 1.

Results of the linear mixed-effects modeling over the full dataset of the VOT ratios show that ASPIRATION, PLACE, GENDER, as well as their two-way interactions (i.e., two-way interactions: AGE:ASPIRATION, AGE:GENDER, ASPIRATION:PLACE) significantly improved the model fits. The full dataset of VOT ratios was further broken into separate subsets according to different plosives given the interactions. Table 3 lists the final full model for each sub-dataset and estimates of the fixed effects, in which *a* for /p/, *b* for /p^h/, *c* for /t/, *d* for /t^h/, *e* for /k/, and *f* for /k^h/.

Table 3: The final model and parametric estimates of VOT ratio for different plosives, *a* for /p/, *b* for /p^h/, *c* for /t/, *d* for /t^h/, *e* for /k/, and *f* for /k^h/.

a. Final model for /p/:
AGE+GENDER+AGE:GENDER+(1|RHYME)

(Baseline for GENDER: FEMALE)

	β	*SE*	*t*	*p*
(Intercept)	10.42	0.69	15.14	<0.001
AGE	0.08	0.03	2.76	<0.01
GENDER	1.09	0.21	5.27	<0.001
AGE: GENDER	-0.18	0.04	-4.75	<0.001

b. Final model for /p^h/: AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	35.15	1.18	29.86	<0.001
AGE	0.12	0.05	2.37	<0.05

c. Final model for /t/: AGE+GENDER+AGE:GENDER+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	9.42	0.33	28.41	<0.001
AGE	0.02	0.02	0.81	n.s.
GENDER	1.47	0.14	10.23	<0.001
AGE: GENDER	-0.10	0.03	-3.87	<0.001

d. Final model for /t^h/: AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	34.49	1.17	29.48	<0.001
AGE	0.13	0.04	3.66	<0.001

e. Final model for /k/: AGE+(1|RHYME)

	β	*SE*	*t*	*p*
(Intercept)	14.88	0.62	24.16	<0.001
AGE	-0.02	0.02	-1.04	n.s.

f. Final model for /k^h/: AGE+GENDER+(1|RHYME) (Baseline for GENDER: FEMALE)

	β	*SE*	*t*	*p*
(Intercept)	35.26	1.00	35.42	<0.001
AGE	0.07	0.04	1.70	n.s.
GENDER	1.13	0.49	2.33	<0.05

Table 3 shows that, AGE showed significant main effects in all aspirated plosives except /k/ (i.e., /p^h/ and /t^h/, see Tables 2*b* and 2*d*), where AGE was the only fixed-factor in the models; for models where there were interactions of AGE and GENDER, a significant simple effect of AGE was only observed for /p/ (see Table 2*a*). In addition, significant intercept differences between genders were observed for /p/ and /t/ as simple effects (see Tables 2*a* and 2*c*) and for /k^h/ as a main effect (see Table 2*f*). As for the raw VOT values, the estimated slope of AGE for each plosive were further calculated for VOT ratios and presented in Table 4. Positive values in the table indicate increasing VOT ratios as a function of age, while negative values indicate declining VOT ratios. * stands for estimated values significantly different from 0 (i.e., no correlation between VOT ratios and age), indicating that the change in VOT ratio is significant.

Table 4: The estimated slope of AGE for each plosive. * stands for estimated values significantly different from 0.

	Female	Male
/p/	0.08 *	-0.10 *
/p^h/	0.12 *	
/t/	0.02	-0.08 *
/t^h/	0.13 *	
/k/	-0.02	
/k^h/	0.07	

As seen from Table 4, significant changes in VOT ratios as a function of AGE can be observed for both bilabial and alveolar plosives, but not for velar plosives. Among those, both aspirated plosives exhibit exclusively increasing trends of VOT ratios across genders (see positive estimated slopes of AGE for /p^h/ and /t^h/ in Table 4), although the magnitudes were small (i.e., <0.5). The directions of change in VOT ratios for unaspirated plosives were gender-dependent. To be specific, rising slopes of change were observed for VOT ratios of /p/ and /t/ produced by female speakers, although the change in /t/ was not significant; significant falling slopes were observed in both plosives produced by male speakers. Regardless of directions, all significant changes in VOT ratios were only in minor magnitudes (i.e., <|0.5|).

4. DISCUSSION

Studies on speech variation have mostly focused on contextual speech variation or individual differences in the acoustic realization of phonetic elements. How age affects the acoustic realization of speech across one's lifetime, however, has long been underestimated in existing models of speech production and perception. The current study thus aims to shed some light on this issue.

By examining both raw VOT and VOT ratio of six plosives in a connected speech corpus of Standard Chinese, the current study observed two phenomena. First, while raw VOT of all plosives gradually developed throughout the middle-aged period, only bilabial and alveolar plosives show significant changes when the influence of speech rate was ruled out as indicated with changes in VOT ratio. Second, the directions of changes, however, are further dependent on speaker gender and manner of articulation. Specifically, while both raw VOT and VOT ratio of aspirated plosives /p^h/ and /t^h/ gradually increase across genders, those for unaspirated plosive /p/ show increasing VOT and VOT ratio only for female speakers, and /p, t/ show decreasing trends for male speakers.

In most previous studies on the effect of age on speech production, stagewise changes have been noticed where older adult speakers show different patterns from those produced by younger adult speakers. Results of the current study therefore further suggest that, such age-related change in plosive VOT is a gradual effect starting from the early stage of adulthood if any, e.g., in one's 30s, instead of an abrupt change before the older stage of life. This finding is thus comparable to the continuous changes of the physiological data [2] as well as observations in vowel production studies for instance [16] as also observed in middle-aged speakers.

Furthermore, the current study also differs from previous studies which claimed consistent trends of change in VOT as the speaker age increases (see e.g., [4, 13, 17, 20] for consistent reduction of VOT in the elderly speakers), but show different directions of change dependent on plosive types and speaker genders. In these previous studies, speakers were asked to produce very limited numbers of stimuli words, so that results were highly influenced by the characteristics of each stimulus. By examining all plosives in Standard Chinese that produced in all possible phonological contexts in connected speech and by taking phonological contexts into considerations in the statistical models, the current study was thus better capable of solving the potential stimuli-dependent problem in previous studies.

In studies which also observed different change directions depending on places of articulation and/or speaker genders, the current results are most aligned with those in [8] where unaspirated voiceless plosives were examined in connected speech. It was found in [8] that, bilabial and alveolar plosives show longer VOT and velar plosives show shorter VOT in older speakers. Although data from both genders of speakers were collected in [8], most of the speakers were females in both old and young groups (i.e., 5 males and 20 females in each group). It is therefore unsurprising that the overall results were in similar trends with those for female speakers in the current study. What is different in the present study is that, when ruling out possible influences from speech rate, the velar plosives only show non-significant changes in VOT ratios. However, the reason why velar plosives exhibited different change trajectories from bilabials and alveolars is in need of further investigations. Results of other studies are of less comparability, for example, in [37], all of the aged subjects suffered from hearing loss to different degrees, who might show very different patterns in the production of plosives.

Previous studies have proposed several explanations for the age-related changes in VOT mainly from the articulatory perspective, none of which, however, accounts for the observed variability of VOT changes in the current study satisfactorily. For example, [8] attributes the observed VOT lengthening to significantly slower rate of articulation in elderly speakers. Explanations along the same line, e.g., slower motoric movements [17] and reduced motoric efficiency [20], have also been proposed. All of these accounts would lead to longer speech sounds and slower speaking rate in general with the increase of age. However, the current study shows that the directions of change in VOT ratio kept comparable to those in raw VOT in most cases (i.e., except for two velar plosives), although the magnitudes of change were suppressed to some extents. This means that speech rate only plays a partial role in the VOT change with age. Moreover, given that the influences from speech rate have been controlled, the directions of VOT change in unaspirated plosives produced by male speakers were deceasing rather than increasing, which contradicts to the predicted direction of VOT change if a slower articulation rate was the only factor.

Other explanations, however, have been proposed to account for the opposite direction of VOT changes. For example, speakers are known with reduced muscle firmness and tissue mass with the increase of age, which further leads to reduction in the subglottal pressure to initiate the vibration of the vocal folds [4]. [8] also argues that, the declining VOT in velar plosives might be due to the decreased lung capacity so that not enough amount of air could be expelled necessary for a plosive. If these were the cases, the shortening of VOT would then be not necessarily limited to plosives in any specific place of articulation.

It is noteworthy that, in a recent study on age-related changes in cue weighting, [38] found that as adult listeners age, they rely less and less on temporal cues (e.g., VOT) to perceive the contrast of voicing, e.g., /b/ vs. /p/ in English, but give more weights to spectral cues (e.g., f0) instead. One of the possible underlying mechanisms has been argued to be the degraded auditory temporal processing in the elderly listeners, so that they are less capable of utilizing small VOT differences as a reliable cue (also see e.g., [26, 32-34]). It is therefore probable that the change of VOT in plosive production might be related to the need of maintaining phonemic contrasts given the degenerated processing ability of temporal cues as speakers age. The possibility of similar adjustments in speech production was also mentioned in [5].

In the case of plosives in Standard Chinese in the current study, as longer VOT is needed to differentiate the contrast in aspiration with the increase of age, speakers might correspondingly lengthen the duration of VOT in their production of plosives. However, to distinguish aspirated plosives from unaspirated plosives, enough differences in VOT must be kept to avoid category overlap. Aspirated plosives are thus more likely to be lengthened (as can be observed from the increase of VOT of aspirated plosives across genders), while the VOT lengthening of unaspirated plosives are rather optimal. This is also in line with the observations in [29] where the VOT values of aspirated plosives show more variability than those of unaspirated plosives in Beijing Mandarin. In the current study, only /p/ produced by female speakers shows significant increase of VOT ratio, while /t/ produced by female speakers does not show significant changes; both /p/ and /t/ produced by male speakers show significant decrease of VOT ratios. Such variability of VOT changes for unaspirated plosives might be related to gender-specific strategies to accommodate the physiological changes due to aging (e.g., [22, 30]) or other sociolinguistic factors as discussed in [38]. Further investigations are needed to verify these possibilities.

Some may argue that the changes in VOT observed in the present study are within-stage changes that occur only in the middle-aged adulthood period, rather than a prelude to the old age (e.g., above 50s). However, this argument is groundless, given that the control of articulatory organs only shows uni-directional decline well into the late old age (e.g., 80 years old) in the physiological studies (e.g., [2]), rather than more complex trajectories. In addition, the current study suggests that, even without data from elderly speakers, the middle-aged speakers already show significant changes in the plosive VOT, which has challenged the received view that there exists a stable stage in the development of speech production in the middle-aged stage. However, it is true that to observe the entire developmental process of speech production across one's lifetime, future studies with data from speakers aged older than 50 are definitely necessary.

One limitation of the present study is that, a cross-sectional experimental design was used, and the statistical analysis was not able to take speaker variation into consideration due to potential collinearity between age and speaker. Although longitudinal studies might show similar direction of VOT change with that in cross-sectional studies (e.g., in [13]), future longitudinal study is in need to better understand the development trajectories of the acoustic properties in speech production across the lifespan.

5. CONCLUSION

Taken together, with evidence from six plosives produced in connected speech in Standard Chinese, the present study shows that even middle-aged speakers show gradually increase VOT in the production of plosives in general. Therefore, at least for phonetic elements such as plosives, the acoustic realization is further developed and in an ongoing, dynamic process of change even in adulthood after a speaker acquires his native language. This thus challenges the traditional perception that speech production is relatively stable throughout younger adulthood and broadens the existing understanding of age-related changes of speech production in general. Such an understanding would hopefully benefit future models of speech production and perception in general.

6. ACKNOWLEDGEMENTS

This study is funded by the National Social Science Foundation of China (Grant No. 19CYY021). Many thanks go to Aijun Li for access to the RASC863 corpus. The assistance from Hongli Liang in data segmentation is greatly appreciated.

REFERENCES

[1] Andres, R. 1976. *The Normality of Aging: The Baltimore Longitudinal Study*. US Department of Health, Education, & Welfare.

[2] Ballard, K. J., Robin, D. A., Woodworth, G., Zimba, L. D. 2001. Age-related changes in motor visuomotor tracking. *Journal of Speech, Language, and Hearing Research* 44, 763–777.

[3] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.

[4] Benjamin, B. J. 1982. Phonological performance in gerontological speech. *Journal of Psycholinguistic Research* 11, 159–167.

[5] Benjamin, B. J. 1997. Speech production of normally aging adults. *Seminars in Speech and Language* 18, 135–141.

[6] Birren, J. E. 1964. *The Psychology of Aging*. Englewood Cliffs, NJ: Prentice-Hall.

[7] Boersma, P., & Weenink, D. 2020. Praat: Doing Phonetics by Computer (6.0.22). http://www.fon.hum.uva.nl/praat/

[8] Bóna, J. 2014. Voice onset time and speakers' age: Data from Hungarian. *Clinical Linguistics and Phonetics* 28, 366–372.

[9] Botwinick, J., Thompson, L. W. 1966. Components of reaction time in relation to age and sex. *The Journal of Genetic Psychology* 108, 175–183.

[10] Burks, S. S., Amerman, J. D. 1979. *Voice Onset Time Characteristics of Geriatric Speakers*. The Annual Convention of the American Speech-Language-Hearing Association.

[11] Chen, Y. 2012. Tonal variation. In: Cohn, A. C., Fougeron, C., Huffman, M. K. (eds), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 103–114.

[12] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27, 207–229.

[13] Decoster, W., Debruyne, F. 1997. Changes in spectral measures and voice-onset time with age: A cross-sectional and a longitudinal study. *Folia Phoniatrica et Logopaedica* 49, 269–280.

[14] Ernestus, M. 2012. Segmental within-speaker variation. In: Cohn, A. C., Fougeron, C., Huffman, M. K. (eds), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 92–102.

[15] Flege, J. E., Eefting, W. 1986. Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* 43, 155–171.

[16] Gahl, S., Baayen, R. H. 2019. Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood. *Journal of Phonetics* 74, 42–54.

[17] Itoh, M., Sasanuma, S., Tatsumi, I. F., Murakami, S., Fukusako, Y., Suzuki, T. 1982. Voice onset time characteristics in apraxia of speech. *Brain and Language* 17, 193–210.

[18] Kaňok, M., Novotný, M. 2019. Effect of age and gender on articulation of voiced and voiceless stop consonants in Czech. *Lekar a Technika – Clinician and Technology* 49, 97–101.

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

[20] Larson, G. W., Hayslip, B., Thomas, K. W. 1992. Changes in voice onset time in young and older men. *Educational Gerontology* 18, 285–297.

[21] Lee, W.-S., Zee, E. 2003. Standard Chinese. *Journal of the International Phonetic Association* 33, 109–112.

[22] Linville, S. E., Rens, J. 2001. Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice* 15, 323–330.

[23] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422.

[24] Neiman, G. S., Klich, R. J. 1983. Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research* 26, 118–123.

[25] Petrosino, L., Colcord, R. D., Kurcz, K. B., Yonker, R. J. 1993. Voice onset time of velar stop productions in aged speakers. *Perceptual and Motor Skills* 76, 83–88.

[26] Pichora-Fuller, M. K., Schneider, B. A., Benson, N. J., Hamstra, S. J., Storzer, E. 2006. Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry. *Journal of the Acoustical Society of America* 119, 1143–1155.

[27] Ptacek, P. H., Sander, E. K., Maloney, W. H., Jackson, C. C. R. 1966. Phonatory and related changes with advanced age. *Journal of Speech and Hearing Research* 9, 353–360.

[28] R Core Team. 2020. R: A Language and Environment for Statistical Computing (3.3.0). https://www.r-project.org

[29] Ran, Q. 2017. Bianyi yu fenhua – Jiaoda yangben shijiao xia de Beijinghua seyin geju [Variation and divergence – The acoustic distribution of stops in Beijing Mandarin based on a large sample]. *Yuyan Wenzi Yingyong（语言文字应用）* 4, 29-38.

[30] Rastatter, M. P., Jacques, R. D. 1990. Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatrica* 42, 312–319.

[31] Sankoff, G. 2018. Language change across the lifespan. *Annual Review of Linguistics* 4, 297–316.

[32] Schneider, B. A., Hamstra, S. J. 1999. Gap detection thresholds as a function of tonal

duration for younger and older listeners. *Journal of the Acoustical Society of America* 106, 371–380.
[33] Snell, K. B. 1997. Age-related changes in temporal gap detection. *Journal of the Acoustical Society of America* 101, 2214–2220.
[34] Snell, K. B., Frisina, D. R. 2000. Relationships among age-related differences in gap detection and word recognition. *Journal of the Acoustical Society of America* 107, 1615–1626.
[35] Sweeting, P. M., Baken, R. J. 1982. Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research* 25, 129–134.
[36] Thomas, E. R. 2013. Sociophonetics. In: Chambers, J. K., Schilling, N. (eds), *The Handbook of Language Variation and Change (2nd edition)*. Chichester: Wiley-Blackwell, 108–127.
[37] Torre, P., Barlow, J. A. 2009. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders* 42, 324–333.
[38] Toscano, J. C., Lansing, C. R. 2019. Age-related changes in temporal and spectral cue weights in speech. *Language and Speech* 62, 61–79.
[39] Wang, T., Li, A. 2003. Lianxu hanyu yuyin shibie yuliaoku de sheji [Design of a connected-speech corpus for speech recognition in Chinese]. Proceedings of the Sixth National Conference of Modern Phonetics.
[40] Zlatin, M. A., Koenigsknecht, R. A. 1976. Developmental of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research* 19, 93–111.

LI Qian PhD She is an associate professor of Institute of Linguistics, Chinese Academy of Social Sciences. Her research interests include the variability in speech production and speech recognition.
E-mail: li_qian@cass.org.cn

［本文原载《中国语音学报》第15辑，2021年］