Effects of Aspiration on Tone Production and Perception in Standard Chinese

Chong Cao, Aijun Li

Chinese Academy of Social Sciences, Beijing 100102, China

caochong2013@126.com,liaj@cass.org.cn

Abstract

Numerous studies reported that the onset fundamental frequency (i.e., onset f0) was usually affected by the voicing characteristics of preceding consonants in speech production. For instance, onset f0 following voiceless stops was usually higher than that following voiced stops. With regards to Standard Chinese, syllable-initial stop consonants could be classified into two groups according to the aspiration contrast, voiceless aspirated and voiceless unaspirated. The aspiration contrast is distinctive and plays an important role in distinguishing lexical meanings. Using acoustic analysis of f0 realization and categorical perception paradigm, the study aims to investigate the effect on the production and perception of lexical tones from consonants' aspiration in Standard Chinese. Production results showed that the onset f0 following aspirated consonants was higher than that following unaspirated syllables. Moreover, the magnitude varied with lexical tones, tone 1 and tone 4 had larger differences in onset f0 than tone 2 and tone 3. Results of perception tests showed that the aspiration contrast enhanced the perceptual salience between high and low tones. Specifically, compared with unaspirated syllables, tones carried by aspirated syllables tended to be perceived as lower tones.

Index Terms: aspiration, f0 perturbation, production, perception

1. Introduction

A large number of studies have documented that consonants in consonant-vowel syllables (/CV/ syllables, hereafter) usually affect the onset fundamental frequency (onset f0, henceforth) of following vowels, leading to local perturbations, which was called "Microprosody" [1] or "CF0" [2-3].

House and Fairbanks (1953) first observed the f0 perturbation due to the obstruent voicing contrast. They found that f0 at the onset was usually higher after a voiceless stop than after a voiced stop [4]. Since then, the phenomenon of f0 perturbation has received growing research attention. Numerous follow-up studies found the similar obstruent-related f0 perturbation effect from non-tonal and tonal languages [5-6].

Although most studies have mainly investigated f0 perturbations caused by obstruent voicing contrasts (i.e., voiceless/voiced), there has been less focus on the influence of consonantal aspiration contrast in onset f0. And research has yet to reach a consensus. Some research reported that the onset f0 following aspirated stops was significantly higher than following unaspirated stops. The results have been demonstrated in several different languages, such as Danish [7], English [5], Korean [8], Thai [9], Cantonese [10] and Taiwanese [11]. Other empirical studies held that the onset f0 was higher after unaspirated stops in contrast with aspirated stops. This assumption was also documented in multiple languages, such as Indian [12], Thai [13], and Wu dialect [14-15]. It was further shown that the magnitude of differences in onset f0 was related to the place of articulation of stops as well as tones. As an example, f0 differences at the onset due to the aspiration contrast were found to be significantly larger after velar stops than after either bilabial or alveolar stops [16].

As for Standard Chinese, a tonal language, uses four tones to distinguish lexical meanings, e.g., ma1 "mother" [Tone 1], ma2 "hemp" [Tone 2], ma3 "horse" [Tone 3], ma4 "scold" [Tone 4]. The aspiration contrast in Standard Chinese is also an important distinctive feature which could distinguish lexical meanings, such as /p^hau/ (run) and /pau/ (full). Despite the fact that prior studies have reported f0 perturbations due to the aspiration voicing contrast in Standard Chinese, studies have yielded conflicting findings. In some studies, the onset f0 following aspirated consonants was higher than following corresponding unaspirated consonants. In another set of studies, the onset f0 following unaspirated consonants was higher than following aspirated consonants. For example, Xu and Xu (2003) adopted a minimal pair of syllables (i.e., /ta/ and /t^ha/), which were embedded into two carrier sentences to investigate the effect of consonants' aspiration on the onset f0 [17]. Results indicated that the onset f0 following unaspirated consonants was much higher than that following aspirated counterparts. In addition, the study found that the f0 perturbation effect was greater for the rising and low tones (i.e., tone 2 and tone 3) than for the high and falling tones (i.e., tone 1 and tone 4). However, this was contrasted by a study of Cao and Zhang (2019), which investigated the effects of aspiration on onset f0 with evidence from affricates that were also paired by the aspirated/unaspirated distinction as stops [18]. Results showed that the onset f0 tended to be higher following aspirated affricates than following unaspirated affricates.

A vast body of academic work has mainly investigated f0 perturbations from the perspective of production. Less attention has been paid to the perceptual effect simultaneously. Yang and Jin (1988) investigated the relationship between stops and tone perception [19]. They found that the articulation manner of stops influenced tone perception: tones carried by aspirated syllables tended to be perceived as tones that have relatively lower onset f0, while tones carried by unaspirated syllables tended to be perceived as tones which have relatively higher onset f0. Cao and Zhang (2019) adopted the similar tone perception paradigm, employed affricates instead, and found the similar perception tendency of lexical tones [18].

The present study further investigated the consonant-related f0 perturbation in Standard Chinese. A primary aim of the current study is to test aspiration's effect on onset f0 in speech production, as well as its interactions with various segments and lexical tones. For this purpose, the current study extended test stimuli by taking the place/manner of articulation of aspirated consonants and all four lexical tones

into account. Another goal of the present study is to examine the effect of consonants' aspiration on lexical tone perception. For this purpose, tone perception experiments (i.e., tone identification paradigm) based on aspirated syllables and unaspirated syllables were conducted.

2. Method

2.1 Production

2.1.1 Participants

20 Chinese speakers (14 females and 6 males) were paid to participate in the experiment. They were all undergraduates from Beijing Language and Culture University. All participants were right-handed, and had no reported history of hearing or speech disorder.

2.1.2 Stimuli

There are six minimal pairs of consonants distinguished by the aspirated/unaspirated contrast in Standard Chinese, 3 pairs of stops (i.e., /p/-/pʰ/, /t/-/tʰ/, and /k/-/kʰ/), and 3 pairs of affricates (i.e., /ts/- /tsʰ/, /tʂ/-/tʂʰ/, and /tɕ/-/tɕʰ/). In order to minimize the interference from other factors, only syllables with isolated vowels were used. Target syllables were selected as follows: first, /CV/ syllables should be combined with four lexical tones to construct meaningful syllables; second, /CV/ syllables should cover above six pairs of consonants as many as possible despite the orthographic restriction. Finally, the minimal pairs /p/-/pʰ/, /t/-/tʰ/ and /tɕ/-/tɕʰ/ combined with the high vowel /i/ were selected as target /CV/ syllables. Apart from target syllables, syllables composed of fricatives, nasals and laterals were serving as fillers. Finally, a total of 52 syllables were selected as the experiment stimuli.

The stimuli were presented in the form of Chinese characters during the process of recording. In order to avoid the influence of word occurrence frequencies in speech production [20], the occurrence frequencies of Chinese characters were balanced. Results of Paired-Sample *t* tests taking the word occurrence frequencies of characters as the dependent variable showed that there were no significant differences between aspirated syllables (M=3.39, SD=0.61) and unaspirated syllables (M = 3.19, SD = 0.63) at a $p < 0.05$ level (t (15) = 1.246, p = 0.232).

2.1.3 F0 annotation, measurement and normalization

The onset and offset f0 of each lexical tone of each target syllable were labeled manually in Praat by a phonetically trained listener. The onset f0 refers to the first full cycle period in the waveform.

Time-normalized f0 was tracked at 11 equidistant points along the tone contours. Manual checking of all obtained f0 data was done to correct any anomaly. To eliminate pitch range differences due to gender, f0 values in Herz scale were converted to specific-speaker z-score transform by applying the following formula: $z = ((x - \mu)/ \sigma)$, where "x" indicates a raw f0 value, "μ" and "σ" refer to the mean and standard deviation of f0 values for each speaker, "z" is the normalized f0 value.

2.1.4 Procedure

Syllables written in Chinese characters were presented on a computer screen, and were produced three times according to a randomized schedule. Prior to the recording, a set of practice session (another 10 syllables) was presented to familiarize the participants with the experiment procedure.

The recording was carried out in a sound-attenuated booth, with a X-RECORDER software in a Dell computer, an M-AUDIO sound card, and a Shure SM 58 microphone. Syllables were recorded with a sampling rate of 22050 Hz and 16-bit accuracy rate in mono channel. Finally, a total of 1920 target syllables were collected.

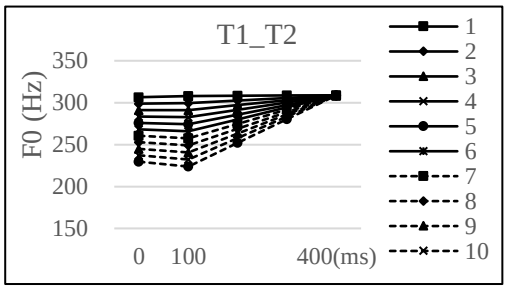
2.2 Perception

2.2.1 Participants

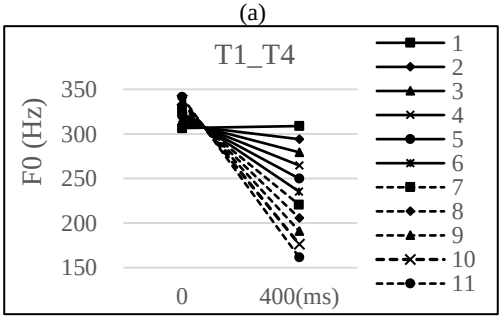
28 Chinese speakers (18 females and 10 males) were paid to participant in the perception experiment. All participants were right-handed, and none of them reported a history of speaking or hearing impairments. They gave informed consent in compliance with a protocol approved by SAIT lab in Beijing Language and Culture University.

2.2.2 Stimuli

Two sets of tone continua, tone 1 _ tone 2 and tone 1 _ tone 4 (T1_T2 and T1_T4, hereafter) based on 3 minimal pairs of syllables (i.e., /pi/-/pʰi/, /ti/-/tʰi/ and /t ɕ i/-/t ɕ ʰi/) were constructed. T1_T2 and T1_T4 continua were adopted in this study because a number of studies have demonstrated that these two tones were perceived categorically by Chinese listeners [21-22].



(a)



(b)

Figure 1: *Schematic diagrams of f0 contours for T1_T2 (a) and T1_T4 continua (b).*

Figure 1(a-b) displays the synthesized f0 contours in each tone continuum. The X-axis represents the normalized duration of each stimulus, the Y-axis represents the f0 value (Hz), and numbers 1-11 in the legend represent the stimulus number in the continuum. Each stimulus's vocalic duration was normalized to 400 msec.

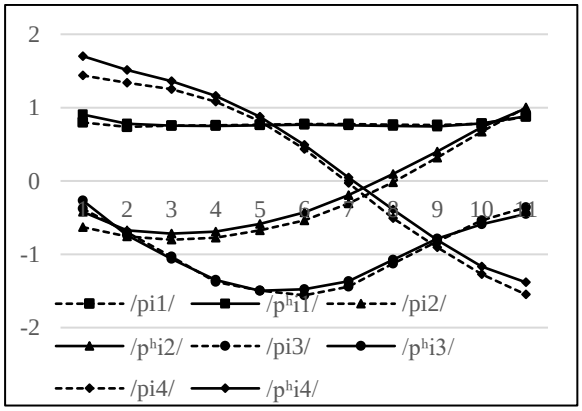
2.2.3 Procedure

Participants were asked to listen to two groups of stimuli separately and to decide the tonal identity by pressing “1” or “2” , and “1” or “4” on the keyboard. For each tone continuum, participants were presented with 3 repetitions of each stimulus in random order. Once a response was collected, the next stimulus was presented automatically. To familiarize participants with the experimental procedure, there was an additional practice session (15 natural syllables were employed as auditory stimuli) before the perception tests.

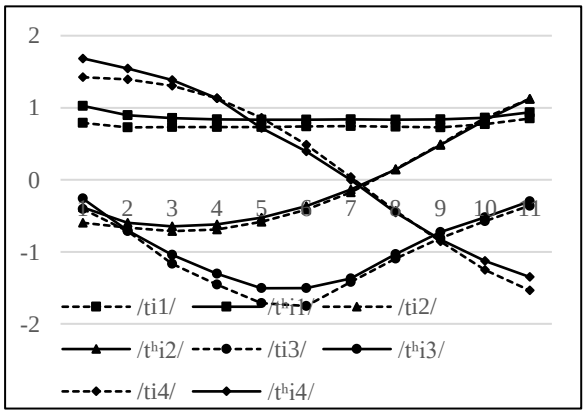
3. Results

3.1 The production results

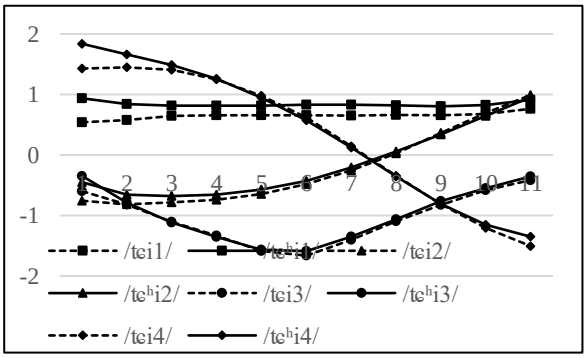
Time normalized f0 contours for 3 pairs of syllables with four lexical tones were shown in figure 2 (a-c), separately for each place of articulation of consonants (i.e., bilabials: /p/-/pʰ/, apicals: /t/-/tʰ/, front palatals: /tɕ/-/tɕʰ/). Each curve represents an average z-score value across 20 speakers. The X-axis represents the normalized time, the Y-axis represents the normalized z-score value of f0. The solid lines indicate aspirated syllables, and the dashed lines indicate unaspirated syllables. Numbers 1-4 represent the four lexical tones, respectively. It can be seen from figure 2(a-c) that, tone 1-tone 4 are realized as a high-level tone, a mid-rising tone, a low-dipping tone, and a high-falling tone, respectively.



(a)



(b)



(c)

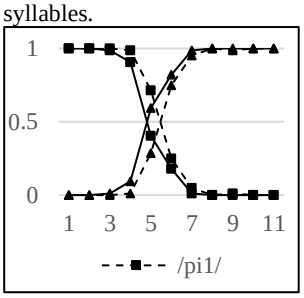
Figure 2: *F0 contours of 4 lexical tones for 3 minimal pairs.*

Looking more closely at figure 2(a-c), we can see that, f0 contours of the same lexical tone carried by aspirated and unaspirated syllables differ in the first several f0 tokens, indicating an impact of aspiration contrast on f0 contours. It is obvious that the onset f0 is substantially higher in the environment of aspiration in comparison to unaspirated environment. The results hold for syllables produced with different place and manner of articulation of consonants as well as different lexical tones.

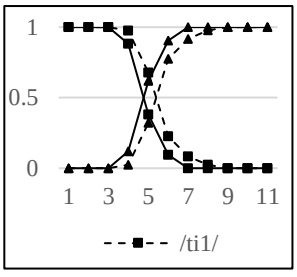
The z-score f0 values were used as the dependent variable in linear multiple regression model. R 3.5.1 was used for statistical analysis (https://www.r-project.org). Results showed that the main effects of aspiration (F (1, 478) = 8.44, p = 0.004) , and tone (F (3, 475) = 1840.1, p < 0.001) were significant.

3.2 The perception results

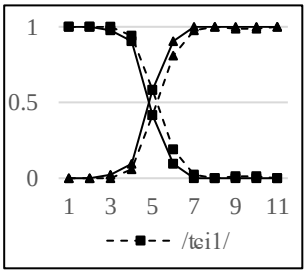
The identification curves of T1_T2 and T1_T4 continua carried by six syllables were plotted in figure 3 (a-f), separated by tone continua (i.e., T1_T2, T1_T4) and minimal pairs (i.e., /p/-/pʰ/, /t/-/tʰ/ and /tɕ/-/tɕʰ/). The X-axis represents the stimulus number in the tone continuum, the Y-axis represents the identification rate of tone 1. The solid lines indicate aspirated syllables, and the dashed lines indicate unaspirated syllables.



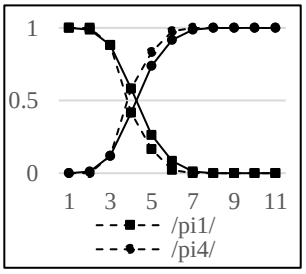
(a)



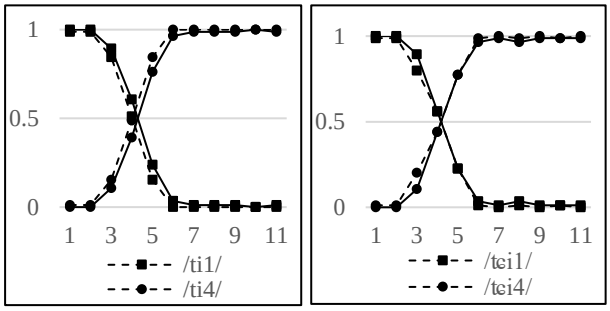
(b)



(c)



(d)



(e) (f)

Figure 3: *Identification curves for T1_T2 and T1_T4.*

A logistic regression between the identification rate and a repeated measures predictor, the step number, was adopted to obtain the mean identification function. The category boundaries of two sets of continua based on 6 syllables are shown in Table 1.

Table 1: *Category boundaries of T1_T2 and T1_T4 continua*.

	Syllables	T1_T2	T1_T4
Unaspirated	/pi/	5.51	3.99
	/ti/	5.49	3.99
	/tɕi/	5.26	4.09
Aspirated	/pʰi/	4.99	4.31
	/tʰi/	4.86	4.32
	/tɕʰi/	4.89	4.29

Statistical analyses taking the category boundaries as the dependent factor were conducted. As to the T1_T2 continua, the main effect of aspiration contrast was significant (F (1,166) = 21.66, $p < 0.005$). The category boundaries of unaspirated syllables were significantly larger than those of aspirated syllables, which indicated that tones carried by unaspirated syllables tended to be perceived as tone 1, a tone with a relatively higher onset, while tones carried by aspirated syllables tended to be perceived as tone 2, a tone with a relatively lower onset. However, there was no significant statistics differences in category boundaries for T1_T4 continua.

4. Discussion and Conclusion

The present study investigated the effects of consonants' aspiration on the production and perception of lexical tones. Results showed that the onset f0 following aspirated consonants was higher than following corresponding unaspirated consonants regardless of lexical tones. The magnitude of differences was greater for tone 1 and tone 4 than for tone 2 and tone 3. Moreover, consonants' aspiration significantly affected the perception of tone 1 and tone 2. Tones based on aspirated syllables were more inclined to be perceived as tone 2; while tones based on unaspirated syllables were more inclined to be perceived as tone 1.

Previous research has demonstrated the articulation differences between voiceless aspirated and voiceless unaspirated consonants in airflow and subglottal pressure [23]. With respect to the airflow, a larger volume of air was released from lungs through the glottis at the oral release phase of aspirated consonants, since the Voice Onset Time (i.e., VOT) of aspirated consonants was, on the average, longer than that of unaspirated consonants. With respect to the subglottal pressure, release generally occurred near the moment of maximal glottal opening for the aspirated consonants, therefore, aspirated consonants exhibited a larger decrease in subglottal pressure compared to that of unaspirated consonants. Prior studies which suggested a higher initial rate of vocal fold vibration of unaspirated consonants, were actually based on the differences in subglottal pressure between aspirated and unaspirated consonants. But this explanation was not supported, indirectly, by Zee's study [10]. This study measured the intensity of voicing of aspirated and unaspirated consonants at the oral release, and found that the subglottal pressure of aspirated stops was lower than that of corresponding unaspirated stops. However, it is interesting to note that, the onset f0 was significantly higher after aspirated stops than after unaspirated stops. Zee's study indicated that a higher onset f0 might be produced even with a lower subglottal pressure.

Two contrary hypotheses on the processing mechanism of f0 perturbation induced by the voicing differences of consonants were proposed. The first suggested that f0 perturbation was produced as an unintended or automatic side effects of differences in vocal fold vibration [24-26]. The hypothesis rested on the assumption that the physiological and/or aerodynamic properties concerning the subglottal/transglottal pressure, the rate of airflow and the vertical tension of vocal cords during the release of oral closure, were possible factors that could be accounted for f0 perturbation. Another influential set of theoretical model emphasized that f0 perturbation might be produced by intended or controlled articulations rather than automatic [27-28]. Speakers deliberately produced f0 differences to exaggerate some kind of perceptual effects or to enhance the consonantal voicing contrasts [29-30]. Perception results of the present study showed that the aspiration contrast did influence the perception of tone 1 and tone 2. More specifically, tones based on aspirated syllables tended to be perceived as tone 2 that has a relatively lower onset f0, while tones based on unaspirated syllables tended to be perceived as tone 1 that has a relatively higher onset f0. From this point of view, f0 perturbations enhanced the perceptual salience between tone 1 and tone 2. However, whether the aspiration-related f0 perturbation in Chinese was automatic or controllable could not be concluded according to the tone perception results. It is necessary to carry out follow-up perception experiments of aspiration consonants to confirm.

There are still some limitations for the present research. An important limitation is that the experimental materials were rather limited, f0 perturbations in a more natural and complicated prosody environment were not investigated. In addition, the auditory stimuli in perception experiments were synthesized based on a linear dimension of f0 (Hz scale) with equal steps, but the nonlinear dimension (logarithmic scale) is more in line with human auditory properties. Further work is needed to look into f0 perturbations in great detail.

5. Acknowledgement

This work is financially supported by National Major Social Sciences Foundation of China (15ZDB103), and China Postdoctoral Science Foundation funded project (2020T130718 and 2020M680811).

6. References

[1] J. M. Hombert, "Consonant types, vowel quality, and tone, " *Tone: A linguistic survey*, 77,112, 1978.

[2] J. Kingston, "Segmental influences on F0: Automatic or controlled?" *Tones and tunes*, vol. 2, pp. 171-210, 2007.

[3] J. P. Kirby, D. R. Ladd, "Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2400-2411, 2016.

[4] A. S. House, G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105-113, 1953.

[5] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 425-441, 2009.

[6] O. Dmitrieva, F. Llanos, A. A. Shultz, A. L. Francis, "Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English," *Journal of Phonetics*, vol. 49, pp. 77-95. 2015.

[7] N. R. Petersen, "The effect of consonant type on fundamental frequency and larynx height in Danish," *Annual Report of the Institute of Phonetics*, University of Copenhagen, vol. 17, pp. 55-86, 1983.

[8] S. A. Jun, "Influence of microprosody on macroprosody: a case of phrase initial strengthening," *UCLA Working Papers in Phonetics*, pp. 97-116, 1996.

[9] W.G. Ewan, "Laryngeal behavior in speech." *PhD diss*, University of California, Berkeley, CA. 1976.

[10] E. Zee, "The effect of aspiration on the F0 of the following vowel in Cantonese," *UCLA Working Papers in Phonetics*, vol. 49, pp. 90-97, 1980.

[11] Y. Lai, C. Huff, J. Sereno, A. Jongman, "The raising effect of aspirated prevocalic consonants on f0 in Taiwanese," *in Proceedings of the 2nd International Conference on East Asian Linguistics*, 2009.

[12] R. Kagaya, H. Hirose, "Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants," *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, vol. 9, pp. 27-46, 1975.

[13] J. Gandour, "Consonant types and tone in Siamese", *Journal of phonetics*, vol. 2, no. 4, pp. 337-350, 1974.

[14] Y. Chen, "How does phonology guide phonetics in segment‐f0 interaction?" *Journal of Phonetics*, vol. 39, no. 4, pp. 612-625, 2011.

[15] F. Shi, "The influence of aspiration on tones." *Journal of Chinese Linguistics*, no. 1, pp. 126-145, 1998.

[16] Y. Lai, "The Effect of Aspiration on the Fundamental Frequency of the Following Tone in Taiwanese." *MA*, University of Kansas, 2004.

[17] C. X. Xu, Y. Xu, "Effects of consonant aspiration on Mandarin tones," *Journal of the International Phonetic Association*, vol. 33, no. 2, pp. 165-181, 2003.

[18] C. Cao, J. S. Zhang, "Effects of affricate's aspiration on realization and perception of lexical tones in Standard Chinese," *The Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 1036-1044, 2019.

[19] Y. F. Yang, L. J. Jin, "Stop Consonants and Tone Perception," *Acta Psychologica Sinica*, no. 3, pp. 236-242, 1988.

[20] Y. Zhao, D. Jurafsky, "The effect of lexical frequency and Lombard reflex on tone hyperarticulation." *Journal of Phonetics*, vol. 37, no. 2, pp. 231-247, 2009.

[21] J. Xi, W. Jiang, L. J. Zhang, H. Shu, "Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course," *Acta Psychological Sinica*, vo. 41, no. 7, pp. 572-579, 2009.

[22] Y. J. Wang, M. J. Li, "The effects of Tone Pattern and Register in Perception of Tone2 and Tone3 in Mandarin," *Acta Psychological Sinica*, vol. 42, pp. 899-908, 2010.

[23] Y. H. Li, "Study on Mandarin Plosives' Airflow and Air-pressure Parameters Based on PAS6600." *in Proceedings of the 4th International Conference on Applied Social Science*, 2014.

[24] J. M. Hombet, J. J. Ohala, W. G. Ewan, "Phonetic explanations for the development of tones," *Language*, vol. 55, pp. 37-58, 1979.

[25] B. Connell, "Tone languages and the universality of intrinsic F0: evidence from Africa," *Journal of Phonetics*, vol. 30, no. 1, pp. 101-129, 2002.

[26] D. H. Whalen, A. Levitt, "The universality of intrinsic F0 of vowels," *Journal of Phonetics*, vol. 23, no. 3, pp. 349-366, 1995.

[27] J. Kingston, "The phonetics and phonology of perceptually motivated articulatory covariation," *Language & Speech*, vol. 35, pp. 99-113, 1992.

[28] J. Kingston, L. D. Randy, "Phonetic knowledge," *Language*, vol. 70, pp. 419-454, 1994.

[29] D. H. Whalen, A. S. Abramson, L. Lisker, M. Mody, "F0 gives voicing information even with unambiguous voice onset times," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2152-2159, 1993.

[30] A. L. Francis, V. Ciocca, V. K. M. Wong, J. K. L.Chan, "Is fundamental frequency a cue to aspiration in initial stops?" The *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2884-2895, 2006.

[This paper was published at ISCSLP 2022]