The effect of duration on categorical perception of Mandarin tones and aspiration of stops

Yan Feng[1,2], *Gang Peng*[1,3]

[1] CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Shenzhen, China

[2] School of Chinese Language and Literature, Central China Normal University, Wuhan, China

[3] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR

ccnuyanny_feng@126.com, gpengjack@gmail.com

Abstract

Whether duration influences the categorical perception (CP) of speech is still a controversial issue. Investigations on the effect of duration of stimuli on the CP has led to opposing results. The present study focused on the effect of duration on the CP of Mandarin lexical tones and aspiration of stops, in which the speech rates were normal for speech production and perception. The results showed that there was no significant duration effect on the CP of lexical tones. However, we found that syllable duration significantly changed the voice onset time (VOT) values of boundary position of aspiration of stops, although the discrimination ability of aspiration was not influenced by syllable duration, indicating that people have a robust perceptual competence to systematically encode alterations of speech production.

Index Terms: duration, tone, aspiration, categorical perception

1. Introduction

Categorical perception (CP) is the ability to perceive continuous acoustic cues in a discrete way [1]. There are many factors affecting the CP in daily communication, such as context and language experience. Whether duration (i.e., speaker rates) influences the degree of the categorical perception of speech is still unclear. Wang and Peng [2] synthesized continua, differing in the duration of stimuli (300 ms vs. 500 ms) and found that the duration had only a limited impact on the CP of Mandarin lexical Tones 1-2. However, Wang, Yang, and Liu [3] selected stimulus durations of 100 ms, 200 ms, and 400 ms, and indicated that a longer duration made the perception of Tones 1-2 more categorical among a Mandarin-speaking older population, while there was no significant duration effect among young people. Chen, Zhu, and Wayland [4] discovered a critical role of stimuli duration in tonal and non-tonal listeners' tone perception, where the duration was from 40 ms to 200 ms. Furthermore, the hemispheric specialization for the perception of formant transition was also shown to be influenced by duration, where the duration of formant transition changed from 40 ms to 80 ms [5]. One possible reason for the controversy may lie in the extreme duration of stimuli used in experiments. It seems that if the stimuli duration is too short or too long, the perceptual competence is weakened. However, given the economy of effort defined by Lindblom [6], it is noteworthy that the extreme speed of articulation is not reached all the time [7], so exploring the effect of natural syllable duration on speech perception is necessary.

Previous research has investigated the duration effect on the perception of voiced and voiceless consonants differing in voice onset time (VOT) based on Indo-European languages. It was concluded that people perceived voiced and voiceless stops in a rate-dependent manner [8, 9]. Furthermore, in studies on duration effect on the perception of fricatives and affricates [10, 11], the duration of fricative noise was found to have a significant effect on perception in Greek and English. However, the classic CP paradigm defined by Liberman et al. [1] was not applied in most of the above studies. What is more, few studies explored whether the perception of aspirated and unaspirated voiceless consonants in Mandarin was influenced by syllable duration. Mandarin has six pairs of aspirated and unaspirated consonants, separated by VOT values, and the VOT values of aspirated and unaspirated consonants are not constant in speech production [12]. That is to say, VOT values of voiceless consonants would increase when the syllable duration is lengthened, and decrease otherwise, especially for aspirated consonants in Mandarin. The unique features of Mandarin phonology offer us a valuable opportunity to check related findings based on Indo-European languages in a cross-linguistic setup.

In contrast to the lexical tone, which is defined mainly by fundamental frequency (F0) patterns, duration is the intrinsic feature of VOT. The present study investigated the effect of duration within natural syllabic rates on the CP of Mandarin tones and aspiration of stops differing in VOT. In this way it was possible to investigate whether duration plays a similar role in both types of phonological contrasts.

2. Methods

2.1. Participants

Sixteen native listeners of Mandarin at the Southern University of Science and Technology, and the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences in Shenzhen (eight female, mean age = 26.2 yr, SD = 2.99 yr) were recruited for this study. All participants were from Northern China and right-handed. None of them had formal musical experience or a history of speech, language, or hearing impairment. A consent form was obtained from each participant with a protocol approved by the Human and Animal Experiment Ethics Committee of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

10.21437/TAL.2018-33

2.2. Materials

The speech samples, including syllable /i/ with high-level tone which means "衣 (clothes)" in Mandarin， syllable /i/ with rising tone that means "姨 (aunt)", and syllables /pa/ and /p'a/ which mean "八 (eight)" and "趴 (grovel)" respectively, were uttered by a native female Mandarin speaker and recorded using Praat [13] with a 44100 Hz sampling rate and 16-bit resolution. All speech stimuli in the tone and aspiration continua were generated by TANDEM-STRAIGHT software [14]. Three tone continua were different in duration, of 250 ms, 375 ms, and 500 ms respectively, as well as three aspiration continua. The F0 of tone continua are shown in Figure 1, and the ranges of VOT values in aspiration continua are presented in Table 1. For tone continua, the differences in duration did not change the step size (8.5 Hz) of the tone stimuli, but a longer duration led to an expanded step size in the aspiration continua. All speech stimuli were presented at 70dB.

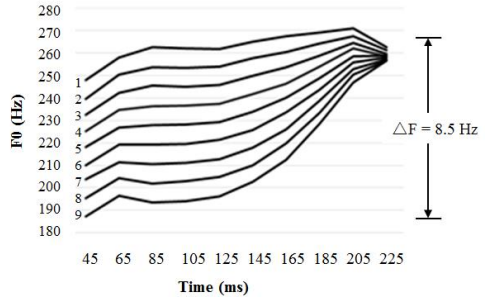


Figure 1: *Tone contours for the stimuli of 250 ms.*

Table 1: *The ranges of VOT value in three aspiration continua.*

Continuum	Range of VOT value (ms)	Step size (ms)
250 ms	5-73	8.5
375 ms	6-106	12.5
500 ms	7-136	16.1

2.3. Procedures

All participants were required to complete both identification and discrimination tests.

2.3.1. Identification test

In the two-alternative forced choices (2AFC) identification test, participants were asked to press key "1" when they heard sound 1 ("衣"/"八") and press key "2" when they heard sound 2 ("姨"/"趴"). There was a practice block for participants to familiarize themselves with the test requirements before a testing block. In the testing block, there were 9 repetitions of each stimulus in each continuum, and all stimuli from each continuum were presented at random. The order of six continua was counterbalanced across participants.

2.3.2. Discrimination test

In the AX discrimination test, participants were instructed to judge whether the pairs they heard, consisting of two stimuli, were different or the same. Key "1" and key "2" represented "the same" and "different" respectively. In each aspiration continuum, there were 25 pairs (1-1, 2-2, 3-3, 4-4, 5-5, 6-6, 7-7, 8-8, 9-9, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 2-1, 3-2, 4-3, 5-4, 6-5, 7-6, 8-7, 9-8). However, only 23 pairs (with each different pair separated by two steps) were included in each tone continuum (1-1, 2-2, 3-3, 4-4, 5-5, 6-6, 7-7, 8-8, 9-9, 1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 3-1, 4-2, 5-3, 6-4, 7-5, 8-6, 9-7), because the one-step tone discrimination task seemed too difficult for participants to complete. There were 5 repetitions of each pair. Prior to the testing block, there was also a practice block with feedback to encourage participants.

2.4. Data analysis

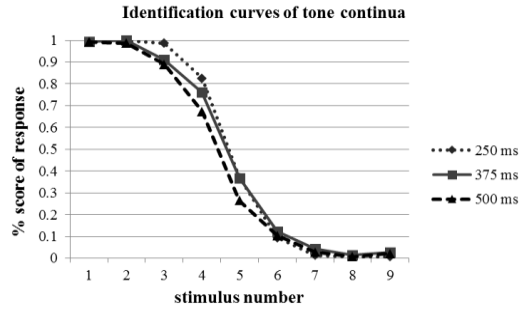
To investigate the impact of duration on the CP of tones and aspiration of stops, four parameters were computed: position of categorical boundary, width of categorical boundary, between-category discrimination accuracy, and within-category accuracy. The identification score was calculated as a percentage of responses. Probit analysis [15] was used to assess the position and width of categorical boundaries, respectively defined as the 50% crossover point and the linear distance between 25% and 75% percentiles. The discrimination accuracy was obtained using the formula proposed in [16]: $P = P("S"|S) \times P(S) + P("D"|D) \times P(D)$.

We divided all pairs into several comparison units, where each unit consisted of four types of pairs (AA, BB, AB, and BA). P (S) and P (D) represented the percentage of AA and BB ("the same"), and the percentage of AB and BA ("different") respectively, both of which were 0.5 in our experiments. P ("S"|S) and P ("D"|D) were defined as the percentage of "the same" and "different" responses to all the same and different pairs. The between-category discrimination accuracy was the average of the discrimination scores of two comparison units straddling the categorical boundary. The within-category accuracy was defined as the average of the discrimination scores of the remaining comparison units.

3. Results

3.1. Identification and discrimination curves

Identification and discrimination curves in different duration conditions are shown in Figure 2. The position of categorical boundary and corresponding F0 range/VOT value of tone and aspiration continua in different duration conditions, and boundary width are presented in Table 2.



(a)

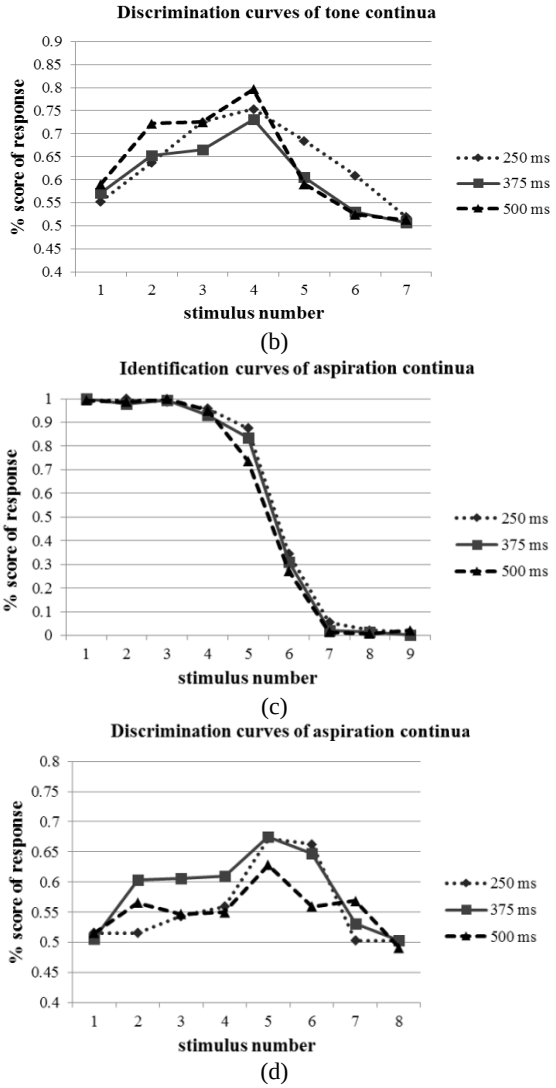


Figure 2: *(a) Identification curves of tones in different duration conditions; (b) Discrimination curves of tones in different duration conditions; (c) Identification curves of aspiration of stops in different duration conditions; (d) Discrimination curves of aspiration of stopes in different duration conditions.*

Table 2: *The mean position and width of categorical boundary of tone and aspiration continua in different duration conditions.*

Continuum	Position	Corresponding F0 range/VOT value	Width
tone_250 ms	4.798	41.783 Hz	0.936
tone_375 ms	4.765	41.503 Hz	1.236
tone_500 ms	4.493	39.191 Hz	1.229
aspiration_250 ms	5.761	45.469 ms	1.182
aspiration_375 ms	5.571	63.138 ms	0.955
aspiration_500 ms	5.498	79.418 ms	1.013

3.2. Position of categorical boundary

To investigate the effect of duration (250 ms, 375 ms, and 500 ms) on the boundary position of tones and aspiration of stops, we conducted one-way ANOVAs on each. There was no significant effect of duration on the boundary position of tone continua ($F(2, 45) = 0.642$, $p = 0.531$), and no significant duration effect on the boundary position of aspiration continua ($F(2, 45) = 0.858$, $p = 0.431$).

One-way ANOVAs were also calculated to discover the impact of duration on the F0 ranges and VOT values corresponding to the boundary position. There was no significant effect of duration on the F0 ranges of tone continua ($F(2, 45) = 0.613$, $p = 0.546$). However, there was a highly significant duration effect on the VOT values separating the VOT distribution of /p/ and /pʻ/ ($F(2, 45) = 71.740$, $p < 0.001$). A Pearson correlation analysis further demonstrated that there was a significant positive correlation between the stimuli duration and the VOT value of the categorical boundary of aspiration of stops ($r = 0.872$, $p < 0.001$), indicating that the VOT value of boundary position of aspiration of stops increased with lengthening stimuli duration, as shown in Table 2.

3.3. Width of categorical boundary

One-way ANOVAs were conducted to explore whether duration would influence the boundary width. For the tone continua, no significant effect of duration was found on the boundary width ($F(2, 45) = 1.530$, $p = 0.228$). For the aspiration continua, there was also no significant duration effect on the boundary width ($F(2, 45) = 0.728$, $p = 0.488$).

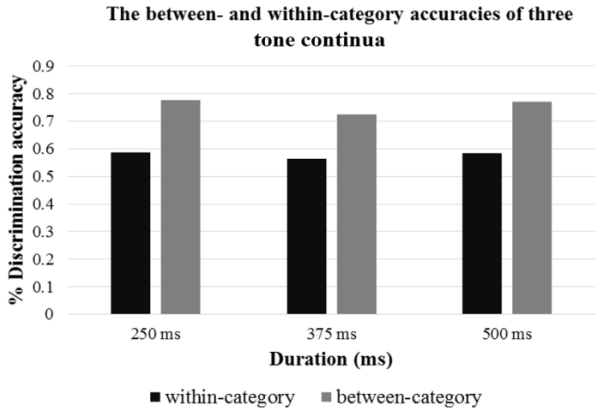
3.4. Discrimination accuracy

The discrimination accuracies of all comparison units of tone and aspiration continua (see detail in Figure 2) were further divided into two categories: between- and within-category, shown in Figure 3.

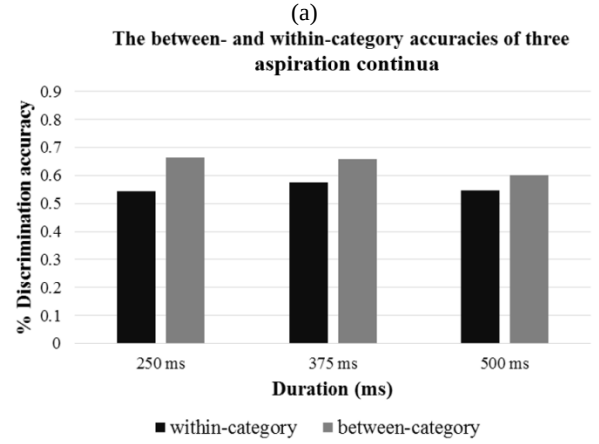
To discover the effect of *duration* (250 ms, 375 ms, and 500 ms) and *category* (between- and within-category) on the discrimination accuracy of tone and aspiration continua, two-way ANOVAs were conducted, with both *duration* and *category* as the within-subject factors. The Greenhouse-Geisser adjustment method was applied when appropriate.

For the tone discrimination, the main effect of *category* was highly significant ($F(1, 15) = 74.530$, $p < 0.001$), indicating that the between-category accuracy was different from the within-category accuracy. As shown in Figure 3(a), the between-category accuracy was always higher than the within-category accuracy, despite the difference in duration. However, there was no significant main effect of *duration* ($F(2, 30) = 1.644$, $p = 0.216$), and no significant *duration* × *category* interaction effect ($F(2, 30) = 0.420$, $p = 0.637$). For the VOT discrimination, the main effect of *category* was

significant ($F(1, 15) = 12.571$, $p = 0.003$), but there was no significant main effect of *duration* ($F(2, 30) = 2.027$, $p = 0.154$). No significant *duration* × *category* interaction effect was found ($F(2, 30) = 1.289$, $p = 0.289$) either.



(a)



(b)

Figure 3. *(a) The between- and within-category accuracy of three tone continua; (b) The between- and within-category accuracy of three aspiration continua.*

4. Discussion

4.1. The effect of duration on the CP of tones

The results from our study support the findings of [2, 3]: there was no significant duration effect on the CP of lexical tones among young adults. This study expanded the range of duration from 250 ms to 500 ms and decreased the interval from 200 ms to 125 ms, which was helpful in conducting more detailed investigations. In contrast, previous research [4] found a strong duration effect on tone perception. Firstly, the range of duration in [4] was from 40 ms to 200 ms, with a step size of 20 ms, whereas the step size was shorter in our study. Secondly, in the study of [3], the effect of duration was mainly observed in the discrimination curves for older listeners whose speech encoding ability and neural flexibility was reported to be weakened because of the aging effect [17, 18]. It is possible that the extremely short duration would have a negative impact on the CP due to physical constraints, such as auditory sensitivity, especially for an aging group.

4.2. The effect of duration on the CP of aspiration of stops

Although the boundary position of the aspiration continua did not change with the stimulus duration, the corresponding VOT value separating the aspirated and unaspirated voiceless consonants increased with lengthening syllable duration. This finding expands on the conclusions of previous studies [8, 9] regarding the CP paradigm. In speech production, VOT values of voiced and voiceless consonants are not fixed, nor are those of aspirated and unaspirated consonants [9, 12, 19, 20], although the latter ones are always longer than the former ones. Our results revealed that people have a robust perceptual competence to systematically encode alterations of speech production. While there is a heated controversy over whether production precedes perception or not [21], shared neural regions were reported to participate both in speech production and perception [22], suggesting there is a relationship between speech production and perception. The results in the present study also support such a relationship, as the perceptual system can accommodate changes in the rate of speech production.

In the present study, duration played a limited role in the boundary width and discrimination accuracy of the aspiration of stops. One possible reason is the ratio of VOT duration and vowel duration in the syllable. Kessinger and Blumstein [20] found that, in speech production, the length of VOT and vowel varied similarly when the syllable duration changed, such that, as the VOT increased, vowel length also increased. However, in our study, the overall length of syllables was the same. Consequently, the vowel length decreased, while the VOT duration increased. That is to say, the length of VOT and vowel changed inversely, which may be inconsistent with findings on production. More investigation is needed to discover the relationships between VOT duration and the CP.

5. Conclusion

In this study, the effect of duration on the CP of tones and aspiration of stops among Mandarin-speaking adults was investigated. We found that the duration of syllable played a limited role in the CP of Mandarin tones. However, our results indicated that the VOT values of the categorical boundary of aspiration of stops increased significantly with a longer duration, while no significant duration effect was found on the boundary width and discrimination accuracy of aspiration of stops, suggesting that people have a robust perceptual competence to systematically encode alterations of speech production.

6. Acknowledgements

This work was supported in part by grants from National Natural Science Foundation of China (NSFC: 11474300), National Social Science Fund of China (13&ZD189).

7. References

[1] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phonemic boundaries," *Journal of Experimental Psychology*, vol. 54, pp. 358-368, 1957.

[2] D. Z. Wang, and G. Peng, "The effects of pitch range and duration on tone categorical perception," *Proceedings of the 3rd International Symposium on Tonal Aspects of Languages*, Nanjing, China. May 26-29, 2012.

[3] Y. X. Wang, X. H. Yang, and C. Liu, "Categorical perception of Mandarin Chinese tones 1-2 and tones 1-4: Effects of aging and signal duration," *Journal of Speech, Language, and Hearing Research*, https://doi.org/10.1044/2017_JSLHR-H-17-0061. 2017.

[4] S. Chen, Y. Q. Zhu, and R. Wayland, "Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions," *PLoS ONE*, vol. 12, no. 7, https://doi.org/10.1371/journal.pone.0180656, 2017.

[5] J. Schwartz, and P. Tallal, "Rate of acoustic change may underlie hemispheric specialization for speech perception," *Science*, vol. 207, no. 21, pp. 1380-1381, 1980.

[6] B. Lindblom, "Economy of speech gestures," in *The production of speech*, edited by P. F. MacNeilage, New York, Springer, pp. 217-245, 1982.

[7] Y. Xu, and X. J. Sun, "Maximum speed of pitch change and how it may relate to speech," *Journal of the Acoustical Society of America*, vol. 111, no. 3, pp. 1399-1413, 2002.

[8] Q. Summerfield, "Articulatory rate and perceptual constancy in phonetic perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, no. 5, pp. 1074-1095, 1981.

[9] J. L. Miller, K. P. Green, and A. Reeves, "Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast," *Phonetica*, vol. 43, pp. 106-115, 1986.

[10] E. Nirgianaki, A. Botinis, and M. Fourakis, "Perception of fricative voice distinctions in Greek," *Proceedings of Fonetik*, Sweden, 12-13 June, 2013.

[11] F. E. Ferrero, G. M. Pelamatti, and K. Vagges, "Continuous and categorical perception of a fricative-affricate continuum," *Journal of Phonetics*, vol. 10, no. 3, pp. 231-244, 1982.

[12] Q. B. Ran, and F. Shi, "VOT analysis of stops in mono-syllables of Standard Chinese," *Nankai Linguistics*, vol. 2, pp. 21-31, 2007.

[13] P. Boersma, and D. Weenink, "Praat: Doing phonetics by computer (Version 5.1.05)," [Computer program]. Retrieved from http://www.praat.org/, 2009.

[14] H. Kawahara, T. Takahashi, M. Morise, and H. Banno, "Development of exploratory research tools based on TANDEM-STRAIGHT," *Proceedings of International Conference on Asia-Pacific Signal and Information Processing Association*. Japan: International Organizing Committee, pp. 111-121, 2009.

[15] D. J. Finney, *Probit Analysis* (3rd ed.). Cambridge University Press, Cambridge, UK, 1971.

[16] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *Journal of Acoustical Society of America*, vol. 120, no. 2, pp. 1063-1074, 2006.

[17] G. M. Bidelman, J. W. Villafuerte, S. Moreno, and C. Alain, "Age-related changes in the subcortical–cortical encoding and categorical perception of speech," *Neurobiology of Aging*, vol. 35, pp. 2526–2540, 2014.

[18] A. Pressaco, K. Jenkins, R. Lieberman, and S. Anderson, "Effects of aging on the encoding of dynamic and static components of speech," *Ear and Hearing*, vol. 36, pp. 352–363, 2015.

[19] Q. B. Ran, CH. N. Liu, and F. Shi, "The categorical perception of aspirated and unaspirated Mandarin stops," *Nankai Linguistics*, vol. 2, pp. 32-39, 2014.

[20] R. H. Kessinger, and S. E. Blumstein, "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," *Journal of Phonetics*, vol. 26, pp. 117-128, 1998.

[21] P. Wong, W. M. Fu, and E. Y. L. Cheung, "Cantonese-Speaking children do not acquire tone perception before tone production—A perceptual and acoustic study of three-year-olds' monosyllabic tones," *Frontier in Psychology*, vol. 8, doi: 10.3389/fpsyg.2017.01450, 2017.

[22] K. Okada, and G. Hickok, "Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI," *Brain & Language*, vol. 98, no. 1, pp. 112-117, 2006.