INITIAL CURVATURE OF PITCH CONTOUR AFFECTS FINE-GRAINED LEXICAL TONE PERCEPTION OF MANDARIN BUT NOT CANTONESE

Yanyuan Ye, Yicheng Rong, Gang Peng

The Hong Kong Polytechnic University, Hong Kong
yanyuan.ye@connect.polyu.hk, rosy.rong@polyu.edu.hk, gang.peng@polyu.edu.hk

ABSTRACT

While the initial curvature (*Wāntóu* in Mandarin), a short curving portion at the onset of a pitch contour in tone languages, has been found inconsequential for distinguishing lexical tones, little is known about the role of this portion in fine-grained tone perception. To address this issue, the current study examined categorisation in 20 Mandarin and 19 Cantonese speakers using two tone continua (/i/ varying from Tone 1 to Tone 2): one continuum was synthesised based on the level tone with initial curvature; the other was based on the one without it. Participants were required to identify lexical tones. Results of Mandarin speakers revealed the initial curvature shifted the labelling boundary towards the rising tone, suggesting the initial curvature may affect fine-grained tone perception. However, this effect was absent in the Cantonese group, indicating the effect of initial curvature might be subject to the complexity of the tone system and different phonetic composition.

Keywords: initial curvature; fine-grained perception; lexical tones; Mandarin and Cantonese.

1. INTRODUCTION

Fundamental frequency (F0) has been widely believed to be the primary cue for tone perception, in which the pitch contour and the register mainly contribute to forming distinctive tones [1], [2]. Although citation tones seem to be stable and well-defined by their pitch contours compared to their sandhi forms, there would still be transitions at both ends of the contour. The transition at the beginning portion of the pitch contour has been described as a prolonged rising [3] and been referred as onset-curving [4] and pre-onset [5]. Given the use of the term ‘onset’ which usually refers to consonants is ambiguous, this portion will be named initial curvature in the current study. The shape of the initial curvature can vary depending on the surrounding phonetic environment.

Previous studies have been trying to define the domain of tone, only several of which focus on the effect of the initial curvature on tone perception. Howie [3] described nine types of F0’s shape depending on different onsets for Mandarin lexical tones and concluded that the domain of tone was the syllabic vowel and the following voiced segment and that the initial curvature of the pitch contour was merely the anticipatory adjustment of the voice.

Lin divided the contour into three parts: the onset-curving section, the tone section, and the end-falling section [4]. He emphasised the essential function of the tone section in distinguishing four tones in Mandarin, which has been supported by his later empirical study [5]. Specifically, he examined the role of length of tone section in tone identification by cutting the sounds into small segments lasting for 80/100ms. The result showed that the longer the segment consisted of the tone section, the higher the identification accuracy would be, while the segments that mainly contained the onset-curving section or the end-falling section would yield non-targeted responses [5]. He thus considered these two sections insignificant for tone classification. Yang [6] conducted another set of perceptual experiments and figured out the perceptual centre of tone was located at the middle parts of the vowel, regardless of any type of onset. Since then, the initial portion of the pitch contour has been considered insignificant for lexical tone perception (see [7] for a review). However, those studies mainly focused on general tone distinguishment and only adopted simple word labelling tasks. Moreover, the materials that previous research used were fragmented or unnaturally synthesised, which might not reflect the authentic pattern of pitch categorisation. Thus, it remains unknown about the role of the initial curvature in fine-grained lexical tone perception.

Although non-tonal languages, such as English and German, do not use lexical tones to differentiate word meanings, changes in pitch may alter the expression of meaning at the sentence level. The transitional F0 movement towards the targeted accented syllable in these languages, i.e., the tonal onglide, has been considered to gain phonological status because it can be used as an acoustic cue to identify pragmatic meanings by German listeners [8], [9]. As the tonal onglide and initial curvature are both transitional adjustments observed at the beginning of pitch contour and thus share articulatory and phonetic similarities, the question that arises is whether the initial curvature would be the perceptual cue like

tonal onglide rather than being merely the phonetic variance.

The current study investigates the effect of initial curvature in fine-grained tone perception in Mandarin and Cantonese by conducting categorical perception experiments, a classical paradigm examining the fine-grained perceptual pattern of categorisation from a micro perspective [10], [11]. We focused on the perception of Tone 1 (high-level tone) and Tone 2 (high-rising tone), as these two tones are acoustically comparable in terms of the overall pitch contour in the two languages we examined [12].

2. METHODOLOGY

2.1. Participants

Twenty Mandarin native speakers (10 males, mean = 21.95 ± 2.98 yrs.) and 19 Cantonese speakers (11 males, mean = 21.88 ± 2.28 yrs.) were recruited. They were all right-handed. None of them majored in music, linguistics, or psychology, nor reported hearing or mental injuries.

2.2. Materials

The Mandarin Tone 1 stimulus *yi1* "clothes" and Tone 2 *yi2* "aunt", and the Cantonese Tone 1 stimulus *ji1* "doctor" and Tone 2 *ji2* "to lean on" were recorded by native female speakers, respectively. As shown in Figure 1d, the Cantonese *ji1* has a small rising at the beginning portion of its pitch contour, while this initial curvature is not observed in the Mandarin *yi1* (Figure 1a), although they share the same overall pitch contour. The four original stimuli were normalised to 300ms and 70dB before further manipulation.

The initial curvature of Cantonese *ji1* was straightened in PRAAT software [13] to form the *ji1* as a pure level tone, without any F0 changing at the beginning portion (see Figure 1e); the Mandarin *yi1* was manipulated with an additional initial rising like the original Cantonese *ji1*, giving rising to *yi1* with initial curvature (Figure 1b).

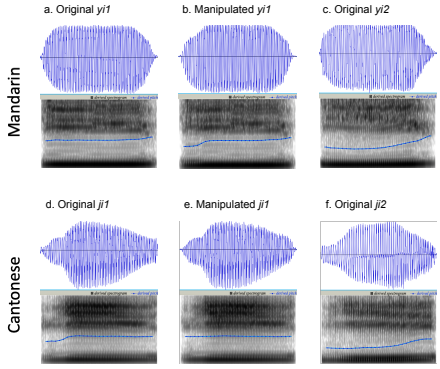

Figure 1: Mandarin tone tokens (the original *yi1* without initial curvature, the manipulated *yi1* with initial curvature, and *yi2*); Cantonese tone tokens (the original *ji1* with initial curvature, the manipulated *ji1* without curvature, and *ji2*)

The four Tone 1 stimuli, together with Cantonese *ji2* and Mandarin *yi2*, were synthesised into four seven-step continua varying from Tone 1 to Tone 2 by applying the pitch-synchronous overlap and add method [14]. As shown in Figure 2, the four continua can be divided into two sets: the Mandarin *yi* with and without initial curvature; the Cantonese *ji* with and without initial curvature. Each continuum has seven stimuli gradually varying from high-level to rising contour with an equal step size (stimulus 1 stands for the typical high-level tone, and stimulus 7 is the rising tone).

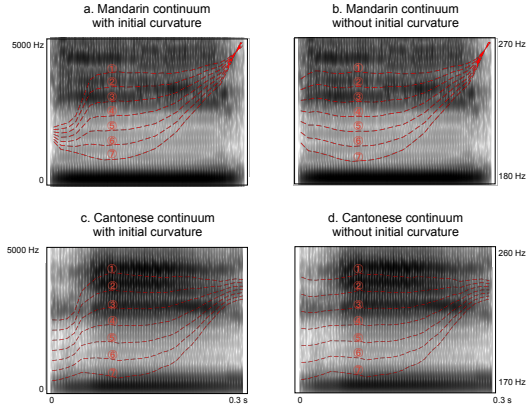

Figure 2: Two Mandarin and two Cantonese continua (the red curves indicate pitch contours along each continuum)

2.3. Procedures

The experiments were carried out remotely via Praat and monitored online via the VOOV platform, which enabled instructing participants as well as ensuring their full concentration during the whole process.

In each language group, there were two identification tasks (two continua: with and without initial curvature) and one rating task. The identification task, a two-alternative-forced choice task, is part of the categorical perception paradigm [10]. Participants were asked to identify the sounds they heard by clicking the mouse. Specifically, the Mandarin speakers need to identify whether they heard was the word "clothes" or "aunt", and the Cantonese participants needed to respond between the word "doctor" and "to lean on". In identification tasks, there were 2 continua * 7 stimuli * 9 repetitions = 126 trials. The order of each task, as well as the presentation of each stimulus, were counterbalanced.

After completing the identification tasks, participants from each language group would rate the naturalness of stimuli 1 and 7 of continua of their native language. Two nonspeech stimuli were taken into the rating task as fillers. On the basis of the templates with Tone 1 and Tone 2, nonspeech

counterparts were synthesized using 64 iteration steps [15]. The rating score ranged from 5 (most natural) to 1 (least natural). The expected rating result was that the naturalness of stimuli with and without initial curvature did not differ in the scores.

2.4. Scoring and data analysis

For each identification task, two key parameters were calculated – namely, the boundary position and the identification slope. First, for each stimulus in one continuum, an identification score, defined as the percentage of responses with which participants identified that stimulus as being Tone 1, was calculated. Next, boundary position, located where the identification score reaches 50%, and identification slope were assessed using a logistic regression equation described as [16]. To determine the influence of initial curvature on boundary position and identification slope, paired samples t-tests were conducted in Mandarin and Cantonese groups, respectively. For the rating task, the naturalness of stimuli with and without initial curvature would be compared in each group using the Wilcoxon signed-ranks test.

3. RESULTS

Figure 3 displays the identification curves for the Mandarin group. The result of paired-sample t-test showed the boundary position of the continuum *yi* with initial curvature (M = 3.70; SD = .36) was significantly larger than that without it (M = 3.46; SD = .31) in the Mandarin group; [$t(19) = -3.11$, $p = .006$, $d = .70$]. In terms of identification slope, there is no significant difference between the continuum *yi* with (M = -1.84; SD = .13) and without initial curvature (M = -1.83; SD = .06); [$t(19) = 0.14$, $p = .891$, $d = .03$].

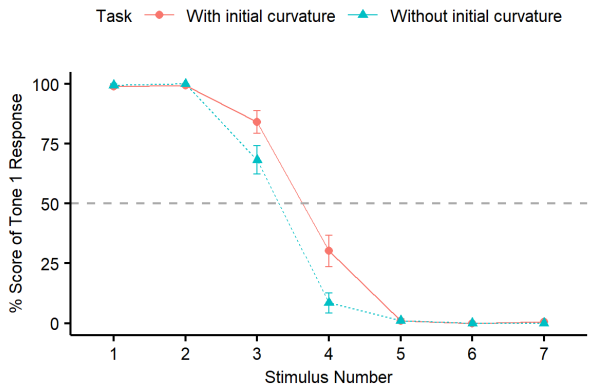

Figure 32: The identification curves of Tone 1 responses for identification tasks with and without initial curvature in Mandarin speakers. Error bars: ± 1 standard error.

As for the Cantonese group, the identification curves are shown in Figure 4. Statistical analyses revealed no significant differences in terms of either boundary position (*ji* with curvature: M = 3.95, SD = .49 vs *ji* without curvature: M = 3.86, SD = .41; $t(18) = 1.35$, $p = .194$, $d = .22$) or identification slope (*ji* with curvature: M = -1.86, SD = .14 vs *ji* without curvature: M = -1.88, SD = .11; $t(18) = .71$, $p = .488$, $d = .08$).

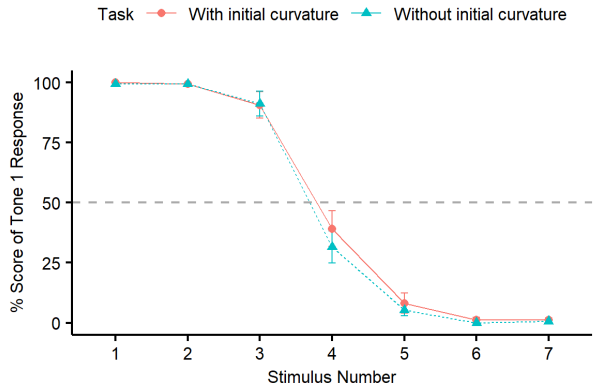

Figure 43: The identification curves of Tone 1 responses for identification tasks with and without initial curvature in Cantonese speakers. Error bars: ± 1 standard error.

Wilcoxon signed-ranks tests were conducted to determine the effect of initial curvature on perceptual naturalness. The results have shown no significant difference between different types of stimuli in either the Mandarin group (*yi* with curvature: M = 4.58, SD = .58 vs. *yi* without curvature: M = 4.61, SD = .43; Z = 102, $p = .472$) or Cantonese group (*ji* with curvature: M = 4.66, SD = .62 vs. *ji* without curvature: M = 4.68, SD = .55; Z = 9.5, $p = .833$), indicating that the naturalness of stimuli would not be affected by the existence or absence of the initial curvature statistically. Moreover, the relatively high mean rating score of each group (almost approaching the highest score of five) has ensured the quality of stimuli in our study.

4. DISCUSSIONS

4.1 The initial curvature contributes to fine-grained lexical tone perception

When Mandarin speakers identified words from the continuum whose base of Tone 1 has additional initial curvature, the identification boundary position would shift towards the rising tone compared to the case when the base had no initial rising. In other words, for Mandarin speakers, if there was initial curvature in the base, there would be more stimuli labelled as the level tone; if the base had no initial rising, the stimuli would be identified more as the rising tone. The initial curvature in Mandarin tones would yield more level-tone responses. However, this effect was not observed in the Cantonese group. There was no significant difference between the two continua in either the identification boundary position or the slope.

Regarding perceptual naturalness, neither the Mandarin nor Cantonese group showed significantly different rating scores between the continua with and without initial rising. The existence or absence of initial curvature would not influence the naturalness of the synthesised stimuli.

Previous studies have argued from the acoustic perspective that the tone domain bears on the syllabic vowel and possible voiced segment following it and that the F0 of the initial voiced consonant or non-syllabic vowel was merely anticipatory movements [3]. Some studies conducted perceptual experiments and concluded that the initial curvature should be excluded from the domain of tone [5], [6]. The results of the present study have suggested that the initial curvature of the pitch contour would affect fine-grained lexical tone perception in some languages. This seemly inconsistency with previous research might result from different designs of experiments and materials: the tasks in previous research were to label the tone class directly. Their primary purpose was to determine which portion contributes to tone distinguishment. The general contour was, without doubt, believed to be foremost in lexical tone perception. However, the issue was far from fully addressed. Having determined the primary perceptual cue did not guarantee the complete irrelevance of other cues. This is what our research cares about. In the present study, we focused on the effect of the short initial portion. Participants were asked to identify between two tones, and the stimuli were from a continuum varying between two tone categories. The stimuli used in the current research were synthesised from the base that contains the entire citation tone and of high naturalness and quality, while those from previous studies were cut and fragmental. To summarise, while the initial curvature has been found inconsequential for distinguishing lexical tones, this portion would affect fine-grained lexical tone perception.

4.2 The effect of initial curvature is subject to the tone system

The effect of initial curvature upon fine-grained lexical tone perception was observed in Mandarin but absent in Cantonese. The phenomenon that needs to be illustrated first is the shifting effect of the initial curvature on the perception boundary of the syllabic vowel *yi* in Mandarin. As introduced, the initial curvature of manipulated *yi* was created by adding a short rising contour at the beginning portion of the vowel [i]. Therefore, we can narrow this question and discuss why changing a small initial portion of F0 would change the boundary position. The shifted boundary position observed among Mandarin speakers, as shown in Figure 3, reflects that the participants were led to yield more level-tone responses, indicating listeners' perception of the pitch direction became less sensitive, and their judgement criteria of the level pitch became looser. One of the possible reasons is that the initial prolonged rising of the F0 contour has reduced the perceptual sensitivity for the rise in the overall pitch pattern. This could also be explained from the perspective of the relationship between acoustics and perception. The presence of the initial curvature gentled the overall slope of the tonal continuum, which in turn affects the perception.

However, this effect observed among Mandarin speakers seems weak in Cantonese speakers. One possible explanation is that the segment bearing the initial F0 curvature of Cantonese is less salient. Different from Mandarin, the phonetic composition of Cantonese stimuli starts with a more fricative onset [j] [17] (see Figure 1 d-e). Another possible reason for the absence of the effect of initial curvature in Cantonese listeners is that the Cantonese tone system is more complex compared to the Mandarin one [12]. Specifically, in Mandarin, Tone 1 and Tone 2 are the only level tone and rising tone, respectively. As more "psychoacoustic space" has been provided, Mandarin listeners would show leniency towards the acoustic changes in the level-tone and rising-tone categories. On the other hand, for Cantonese listeners, the category of Tone 2 is compressed due to the existence of Tone 5, the other rising tone in the Cantonese tone system. Likewise, the other two level tones in Cantonese (Tone 3 and Tone 6) will press the category of Tone 1 into a smaller psychoacoustic space, making this Tone 1 category less flexible and changeable relative to that of Mandarin. In short, the different complexity of the tone system might alter the likelihood of adjusting the perceptual categories of tones: the more complex and denser the system, the less it is likely to be affected.

5. CONCLUSIONS

The current study investigated the influence of initial curvature of pitch contour on the fine-grained lexical tone perception among speakers of Mandarin and Cantonese. Our findings have revealed that the presence of the initial curvature would yield more level tone responses in Mandarin listeners by reducing their sensitivity to pitch direction. This effect was not observed in Cantonese, indicating that the effect of initial curvature might be subject to the complexity of the tone system and different phonetic composition.

6. ACKNOWLEDGEMENTS

This research was partly supported by a fellowship award from the Research Grants Council of the Hong Kong SAR, China (Project No. PolyU/RFS2122-5H01).

7. REFERENCES

[1] Y. R. Chao, 'Mandarin primer', in *Mandarin Primer*, Harvard University Press, 2013.
[2] Y. Xu, 'Contextual tonal variations in Mandarin', *J. Phon.*, vol. 25, no. 1, pp. 61–83, Jan. 1997, doi: 10.1006/jpho.1996.0034.
[3] J. M. Howie, 'On the domain of tone in Mandarin', *Phonetica*, vol. 30, no. 3, pp. 129–148, 1974.
[4] M.-C. Lin, 'The pitch indicator and the pitch characteristics of tones in standard Chinese', *China ACTA Acust.*, vol. 1, pp. 8–15, 1965.
[5] M.-C. Lin, 'A perceptual study on the domain of tones in Beijing Mandarin', *China ACTA Acust.*, vol. 20, no. 6, pp. 437–445, 1995.
[6] S. Yang, 'A preliminary study on the perceptual center of tones in Standard Chineseb', *Acta Psychol. Sin.*, vol. 3, pp. 247–253, 1992.
[7] Y. Song, 'Hanyu Shengdiao Shiyan Yanjiu Huiwang [A review of Mandarin tone studies]', *Yuwen Yanjiu*, no. 1, pp. 41–45, 2006.
[8] D. Crystal, 'Prosodic Systems and Intonation in English. Cambridge Studies in Linguistics 1.', 1969.
[9] S. Ritter and M. Grice, 'The Role of Tonal Onglides in German Nuclear Pitch Accents', *Lang. Speech*, vol. 58, no. 1, pp. 114–128, Mar. 2015, doi: 10.1177/0023830914565688.
[10] B. H. Repp, 'Categorical Perception: Issues, Methods, Findings', in *Speech and Language*, vol. 10, Elsevier, 1984, pp. 243–335. doi: 10.1016/B978-0-12-608610-2.50012-1.
[11] McMurray B., 'The Myth of Categorical Perception'. PsyArXiv, Feb. 21, 2022. doi: 10.31234/osf.io/dq7ej.
[12] G. Peng, 'Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese', *J. Chin. Linguist.*, vol. 34, no. 1, p. 134, 2006.
[13] P. Boersma and D. Weenink, *Praat: doing phonetics by computer (Version 6.1. 09)[Computer program]. Retrieved May 1, 2019*. 2019.
[14] E. Moulines and J. Laroche, 'Non-parametric techniques for pitch-scale and time-scale modification of speech', *Speech Commun.*, vol. 16, no. 2, pp. 175–205, Feb. 1995, doi: 10.1016/0167-6393(94)00054-E.
[15] J. Swaminathan, A. Krishnan, J. T. Gandour, and Y. Xu, 'Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem', *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 281–287, 2007.
[16] Y. Xu, J. T. Gandour, and A. L. Francis, 'Effects of language experience and stimulus complexity on the categorical perception of pitch direction', J. Acoust. Soc. Am., vol. 120, no. 2, pp. 1063-1074, Aug. 2006, doi: 10.1121/1.2213572.
[17] X. Li, J. Huang, Q. Shi, Y. Mai, and D. Chen, 'Guangzhou Fangyan Yanjiu [Research on Guangzhou Cantonese]', 1995.