RELATING ACOUSTIC PROPERTIES OF MANDARIN TONES TO PERCEPTUAL CUE WEIGHTS

Keith K.W. Leung and Yue Wang

Language and Brain Lab, Department of Linguistics, Simon Fraser University, Canada
kwl23@sfu.ca, yuew@sfu.ca

ABSTRACT

This study explores the production-perception relation of Mandarin tones by predicting native Mandarin speakers' tone perceptual cue weights using acoustic features of their tone productions. Participants produced the Mandarin syllable /i/ with four tones and performed a speeded discrimination task. A fundamental frequency (F0) mean was computed for each production and a parabola was fitted to each tone contour to determine F0 slope and curvature. A tone height X tone direction perceptual tone space and individual cue weights were obtained using the Individual Difference Scaling (INDSCAL) analysis. The F0 mean X F0 slope tone space provided the closest match to the perceptual tone space. The multiple linear regression with tone-to-tone distances of each acoustic feature as predictors of the corresponding cue weights did not find a significant production-perception relationship. Findings suggest that more fine-grained acoustic correlates and dimensions are needed to establish patterns of cue weights in tone production-perception link.

Keywords: tone perception, tone acoustics, perception-production relationship

1. INTRODUCTION

A link between production and perception is predicted by speech perception theories [5, 8], and supported by empirical findings [1, 6, 15]. However, these theories and findings are based on segments: e.g. plosives [13], fricatives [15] and vowels [6]. This study aims to extend the study of production-perception link to the suprasegmental level using Mandarin tones as a test case.

Mandarin is a lexical tone language with four tone categories differing in pitch height and contour. Tone 1 (T1) has a high-level contour. Tone 2 (T2) has a mid-high-rising contour. Tone 3 (T3) is a low dipping tone and Tone 4 (T4) is a high-falling tone [4]. Each Mandarin syllable carries a tone and the lexical meaning of the syllable changes if the tone varies. It has been established that Mandarin tone perception involves two main perceptual dimensions: Tone height and tone direction [9]. Cue weights given to these dimensions have been found to vary cross-linguistically and native Mandarin listeners give stronger weights to tone direction than non-tonal native listeners [9, 10]. In Mandarin tone production, it is generally assumed that F0 mean and F0 slope are the two acoustic correlates to tone height and direction, respectively, based on the tone category distributions on the Mandarin tone space [3, 16].

However, F0 slope as a linear function does not represent the more detailed, within-tone acoustic cues (e.g. temporal location of turning point (TP) and F0 decrease from tone onset to TP (ΔF0)) that have been shown to characterize both Mandarin tone perception and production, especially T2 and T3 [14]. In Gandour's [9] phenomenal study about tone perceptual dimensions, the dimension of tone direction is not only represented by linear, but also by rise-fall and fall-rise contours. Therefore, a tone modelling method should reflect within-tone F0 changes. In fact, tone resynthesis and modelling have employed at least a second-order polynomial function [19, 22, 24]. It is possible that F0 curvature [22] modelled by a polynomial function, instead of F0 slope, is a better acoustic correlate of the perceptual dimension of tone direction. This approach of modelling tones and relating it to perception has not been widely used [11].

This study examines Mandarin tone production-perception link by exploring the relationship between perceptual cue weights (i.e. tone height and tone direction) and tone contrasts made with the corresponding acoustic cues that include curvature (F0 mean, F0 slope, and curvature). It is expected that, if a production-perception link exists for Mandarin tones, the individual differences in cue weightings can be predicted by acoustic cues used for contrasting tone categories [3, 6, 21]. For example, the individuals who generally produce a greater tone contrast using F0 mean than other individuals should also weight tone height more strongly than the others.

2. METHOD

2.1. Participants

Ten native Mandarin participants (7 female) recruited at Simon Fraser University participated in this study.

All participants completed the production and perception tasks.

2.2. Production task

2.2.1 Stimuli

Three syllables /i/, /u/ and /a/ with four Mandarin tones, resulting in 12 real words, were used in this task. The four tone words with /i/ were used for acoustic analysis while other syllables were fillers. The syllables in each tone were presented in Chinese characters and in phonetic symbols.

2.2.2 Procedures

In a self-paced production task, the words and phonetic symbols were presented one at a time in the centre of a computer screen. Their productions were recorded digitally in a sound-attenuating booth at a sampling rate of 48 kHz. They were instructed to say each word naturally. The /i/ words occurred 6 times per tone and the other two words occurred 3 times per tone. There were 48 trials (1 syllable (/i/) X 4 tones X 6 repetitions + 2 syllables (/u/ and /ma/) X 4 tones X 3 repetitions).

2.3. Perception task

2.3.1. Stimuli

The stimuli were the syllable /i/ with four Mandarin tones, produced by a male native Mandarin speaker and recorded in a sound attenuated booth. Each tone target was repeated 3 times and 2 repetitions with similar duration (388-421ms) were selected. The amplitude was normalized at 60dB.

2.3.2. Task

Participants performed an AX discrimination task (following [3]) using Paradigm 2.3 [18]. Stimuli were presented binaurally and participants were asked to indicate whether the tone pairs had the same or different tones by pressing the left or right button on the keyboard. The buttons were labelled with SAME or DIFFERENT and the button assignment was counterbalanced. Each trial contained a pair of stimuli. The two repetitions of each tone combined to form the same pairs, whereas only the first repetition was used to form the different pairs. For all possible pairings, two orders of presentation were included (e.g. for same pairs, repetition 1 vs 2, repetition 2 vs 1; for different pairs, tone 3 vs 4 and tone 4 vs 3). To make sure that the "same" and "different" trials had equal probability of occurrence, the 4 "same" pairs were presented 12 times and the 6 "different" pairs were presented 8 times. Therefore, there were a total of 192 trials (4 same pairs x 2 orders x 12 times + 6 different pairs x 2 orders x 8 times) divided equally in 2 blocks (each contained 96 trials). The order of trials was randomized in each block. The interstimulus interval (ISI) was 500ms, following previous tone studies [3, 10]. Participants were given a maximum 1.5s response interval after each pair and were asked to respond as quickly as possible. Reaction time was calculated from the end of the second stimulus presentation.

2.4. Analysis

2.4.1. Production task

The average F0 (Hz) and the F0 values at each 10% interval of each tone contour produced by each participant [23] were measured by Praat [2] using the autocorrelation method, with the F0 range set to 50–400 Hz. To normalize for inter-speaker pitch range differences, each frequency value was then converted from Hz to a logarithm-based T value using the following formula [20, 23]:

(1) $T = (\log x - \log L)/ (\log H - \log L)$ X 5

where x is F0 value in Hz at any given point, L and H are the minimum and maximum F0 produced by the speaker. T has a range of 0 to 5, corresponding to the pitch scale for lexical tones developed by [4].

Polynomial fits were used to estimate slope using (2) and curvature using (3) based on the normalized F0 values at the 11 points (following [23]) along the tone contour using the following functions [11, 22]:

(2) $F(t) = mt + k$

(3) $F(t) = at2 + bt + c$

where t represents the time elapsed from the tone onset, obtained by multiplying the total duration of each tone contour by the relative location of each time point along the tone contour. The linear coefficient of the linear function (m) and the quadratic coefficient of the quadratic function (a) represent slope and curvature respectively. Positive and negative slope values reflect rising (T2) and falling contours (T4), whereas positive and negative curvature values represent best fit parabolic lines with an upward (T2 and T3) and a downward opening (T4). A more curved contour should yield a larger curvature value than a less curved contour. Twenty-seven productions (out of 240) with a creaky voice were removed since the extremely low F0 would distort tone normalization and modelling results.

2.4.1. Perception task

Perceptual dimensions were determined by INDSCAL analysis (following [3]). The input consisted of 26 data matrices. Each matrix contains

the distance estimate for each tone pair (4 tones x 4 tones) per speaker. For obtaining distance estimates, the reaction time data were converted to dissimilarity scales, i.e. the inverse of reaction time: 1/RT, assuming listeners spend longer time to discriminate between two sounds if they are more alike in a perceptual space. Data points greater than 2 SD from the mean were removed.

The output of INDSCAL contained a group tone stimulus space for all listeners in normalized distance. The output also includes each listener's weightings of each dimension.

3. RESULTS

3.1. Production results

Figure 1: Four Mandarin tone contours produced by one speaker

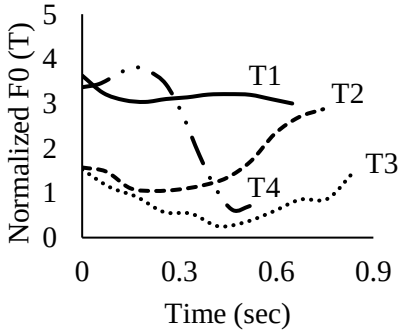


Table 1: Mean acoustic values of four Mandarin tones

Tone	Mean	Slope	Curvature
T1	3.20	0.70	2.81
T2	2.40	6.44	15.72
T3	1.18	0.32	19.14
T4	2.84	-12.05	-49.49

The F0 mean, slope and curvature were measured from each speaker's productions. Averaging across all speakers (Table 1), as expected, T1 as a high tone, had the highest F0 mean than other tones, followed by T4, T2 and T3 (Refer to Figure 1 for tone contour examples). T2 and T4 showed a positive and negative F0 slope values respectively and T1 and T3 had F0 slope values near zero. As a level tone, T1 was the least curved as shown by the mean F0 curvature value closest to zero. T2 and T3 were a parabola opening upward which was consistent with their rising and dipping contour, respectively. The falling contour of T4 yielded a negative curvature value representing a parabola opening downward.

3.2. Perception results

In the discrimination task, participants achieved the mean accuracy rate of 97% and mean reaction time of 476 msec. The 2-dimensional INDSCAL model accounted for 85.7% of the variance (the analysis was restricted to a 2-dimensional solution – half of the category number (i.e. 4 tones/2 = 2)). As shown in the group stimulus space (Figure 2), the positions of tone categories were consistent with [3]. The interpretation of the dimensions followed [3]. Dimension 1 and 2 corresponded to tone height and tone direction respectively.

Figure 2: Group stimulus space obtained from the INDSCAL analysis.

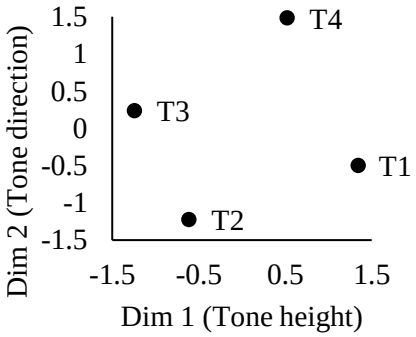
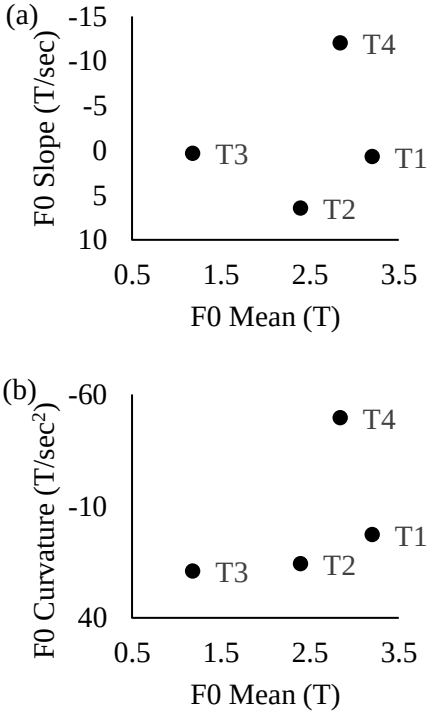


Figure 3: Tone production space in terms of (a) F0 Mean and F0 Slope, and (b) F0 Mean and F0 Curvature.



To examine the acoustic correlates of the perceptual dimensions obtained from INDSCAL, the acoustic

values were plotted on 2 two-dimensional tone spaces, with F0 mean on the *x*-axis and either F0 slope or F0 curvature on the *y*-axis. The tone space with F0 mean and slope as dimensions shared the closest tone category locations with the INDSCAL perceptual tone space (Figure 3a). The F0 mean x F0 curvature plot showed T1, T2 and T4 locations comparable to the perceptual space, but T3 was placed on a lower position than T1 and T2 (Figure 3b). As a level tone, it was not surprising that T1 had a smaller curvature value than T3. T3 involved a slightly more curved shape than T2 in general. In other words, T3 had a more rapid change in F0 than T2. However, this difference was not reflected in terms of F0 slope or tone direction. As a result, F0 mean and slope were the cues that correspond to tone height and direction in perception based on this visual inspection.

3.3. Production-perception relationship

The production-perception relationship was examined using a multiple linear regression analysis with all predictors entered to the model at the same time. The perception measures were each participant's weights given to each perceptual dimension returned by INDSCAL. Production measures were *d*' scores obtained for each of the three acoustic cues and each tone pair using the following formula [6, 21]:

(4) $d'_{axy} = (m_{ax} - m_{ay}) / [(s_{ax}^2 + s_{ay}^2)/2]$

where *a* is an acoustic cue, *x* and *y* are two tone categories, m_{ax}, s_{ax}, m_{ay} and s_{ay} are the mean and the standard deviation of the acoustic cue *a* of tone *x* and *y*. A higher *d*' value represents that the difference of a particular acoustic cue between two tone categories is more detectable.

The *d*' scores of F0 mean (the acoustic correlate of F0 height) of all six tone pairs (i.e. T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) were used to predict the weights given to tone height. As possible acoustic correlates of F0 direction, the *d*' scores of F0 slope or curvature of all six tone pairs were used to predict the weights given to tone direction. The models did not yield any significant result (Tone height-F0 mean: $F(6,3) = 0.568$, $p = .745$; Tone direction-F0 slope: $F(6,3) = 0.630$, $p = .712$; Tone direction-F0 curvature: $F(6,3) = 1.072$, $p = .519$; level of significance: $p = .05$).

4. DISCUSSION

The present study shows that Mandarin tone production and perception are related at group level but not at individual level. Tone height and direction are the two perceptual cues for native Mandarin individuals, consistent with previous findings from similar studies [3, 7, 9, 10]. The acoustic correlates of these dimensions are best interpreted as F0 mean and F0 slope, rather than F0 curvature, demonstrated by a better match of production and perceptual tone spaces. Therefore, general Mandarin tone categorization involves overall tone direction represented by slope (rising or falling) and curvature plays a less crucial role.

The lack of a significant production-perception relationship in the multiple linear regression analysis indicates that the contrast of a particular acoustic cue between tone categories may not be related to the strength of cue weights given to the corresponding perceptual dimension at individual level. Given the small sample size in this study, the non-significant results may be due to the lack of statistical power. On the other hand, previous studies have related speech production to either perceptual cue weights [6] or the acoustic properties that modulate perception [1, 15]. This study suggests that the first approach may not be the manner that tone production and perception are related. However, it is possible that a tone production-perception relationship may instead exist between the acoustic characteristics of tone productions and the same properties that modulate tone perception. This possibility needs to be examined in future studies using an identification task to measure perception responses.

In addition, more detailed tone acoustic cues may be needed to establish a perception-production relationship [15]. The two main perceptual dimensions cannot reflect these details since F0 direction corresponds better to F0 slope than to F0 curvature. As a result, this perceptual dimension does not reflect the within-tone acoustic properties like ΔF0 and TP, as mentioned previously. Therefore, the current findings indicate that further studies should consider specific acoustic parameters in order to examine the Mandarin tone perception-production relationship.

To conclude, this study shows an alignment of perception and production tone spaces at group level when the dimensions are interpreted as F0 mean and slope: the more general representations of tone contours. The non-significant production-perception link from the multiple linear regression suggests that more fine-grained acoustic cues that modulate perception may need to be examined to establish a tone production-perception link.

5. ACKNOWLEDGEMENTS

We thank Prof. Murray Munro and Dr. Henny Yeung for their valuable inputs. This research was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada: 2017-05978.

6. REFERENCES

[1] Beddor, P. S. 2015. The relation between language users' perception and production repertoires. *Proc 18th ICPhS* Glasgow, 171–204.

[2] Boersma, P., Weenink, D. 2018. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.43. Retrieved from http://www.praat.org/

[3] Chandrasekaran, B., Sampath, P. D., Wong, P. C. M. 2010. Individual variability in cue-weighting and lexical tone learning. *J. Acoust. Soc. Am.* 128, 456–465.

[4] Chao, Y. R. 1967. *Mandarin primer: an intensive course in spoken Chinese* Cambridge, MA: Harvard University Press.

[5] Diehl, R. L., Lotto, A. J., Holt, L. L. 2004. Speech Perception. *Annu. Rev. Psychol.* 55, 149–179.

[6] Fox, R. A. 1982. Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica* **39**, 1–22.

[7] Francis, A. L., Ciocca, V., Ma, L., Fenn, K. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *J. Phon.* 36, 268–294.

[8] Galantucci, B., Fowler, C. A., Turvey, M. T. 2006. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.

[9] Gandour, J. T. 1983. Tone perception in far eastern-languages. *J. Phon.* 11, 149–175.

[10] Guion, S. G., Pederson, E. 2007. Investigating the role of attention in phonetic learning. In: Bohn, O.-S., Munro, M. J. (eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 57–77.

[11] Leung, K. K. W., Wang, Y. 2018. The relation between production and perception of Mandarin tone. *J. Acoust. Soc. Am.* 144, 1721–1721.

[13] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychol. Rev.* 74, 431–461.

[14] Moore, C. B., Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *J. Acoust. Soc. Am.* 102, 1864–1877.

[15] Newman, R. S. 2003. Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *J. Acoust. Soc. Am.* 113, 2850–2860.

[16] Peng, G. 2006. Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *J. Chinese Linguist.* 34, 134–154.

[17] Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., Wang, W. S. Y. 2010. The influence of language experience on categorical perception of pitch contours. *J. Phon.*. 38, 616–624.

[18] Perception Research Systems 2007. *Paradigm Stimulus Presentation*, Retrieved from http://www.paradigmexperiments.com

[19] Prom-on, S., Xu, Y., Thipakorn, B. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *J. Acoust. Soc. Am.* 125, 405–424.

[20] Rose, P. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Commun*. 6, 343–352.

[21] Shultz, A. A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing, *J. Acoust. Soc. Am.* 132, EL95-EL101.

[22] Tupper, P., Leung, K. K. W., Wang, Y., Jongman, A., Sereno, J. A. 2018. Identifying the distinctive acoustic cues of Mandarin tones. *J. Acoust. Soc. Am.* 144, 1725–1725.

[23] Wang, Y., Jongman, A., Sereno, J. A. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am.* 113, 1033–1043.

[24] Zhao, T. C., Kuhl, P. K. 2015. Effect of musical experience on learning lexical tone categories. *J. Acoust. Soc. Am.* 137, 1452–1463.