HOW PERCEPTIBLE IS THE DIFFERENCE BETWEEN TONE 3 AND TONE 4 MANDARIN?

Irene Vogel*, Angeliki Athanasopoulou+, Chao Han*, Yue Yan*

* University of Delaware, + University of Calgary
hanchao@udel.edu, ivogel@udel.edu, angeliki.athanasopou@ucalary.ca, yueyan@udel.edu

ABSTRACT

Mandarin tones are "prescriptively" distinguished by four pitch (F0) patterns. Speakers may not always produce these distinctions clearly, however, resulting in perceptual confusability. We investigate Tone 3 (T3, dipping) vs. Tone 4 (T4, falling), which share an initial falling contour, with monosyllabic words extracted from a large corpus of connected speech. Mandarin listeners (N=17) hear the words, and select one of four characters representing words differing only in tone. We observe considerable confusion, but a higher overall accuracy for T4. A d' analysis confirms a T4 bias, and yields more similar confusion rates for both tones without the bias. The T3 and T4 confusability patterns are additionally compared to those involving other tone combinations, found to be less readily confused. Acoustic measurements of the same corpus suggest that the main source of confusability is the relatively smaller role of F0 in the T3 vs. T4 distinction (similar initial fall), compared with other distinctions.

Keywords: Mandarin, tones, tone perception, tone production.

1. INTRODUCTION

Mandarin Chinese is described as having four contrastive tones, differing in their pitch properties: Tone 1 (T1=high), Tone 2 (T2=rising), Tone 3 (T3=dipping), Tone 4 (T4=falling). These are, however, prescriptive characterizations of the tones, and in more natural connected speech, it has been observed both that the tones are not always fully articulated [11, 13, 14], and that they may be perceptually confused [8, 9, 15, 16].

Previous studies typically base their descriptions of the tones when said in isolation [3, 10], or in single syllable words [4, 6, 7]. Additional properties may also affect the tones if words are produced in a list, at the end of a carrier sentence, or in a position in which they are also focused. Moreover, tone perception studies do not necessarily test the same types of items used in production studies, so it is not possible to directly compare the perceptual distinguishability of the tones and their production properties.

In this study, we investigate the perceptibility of the Mandarin tones, focusing primarily on T3 and T4, both of which include an initial falling contour. In connected speech, moreover, the second (rising) part of T3 tends to be reduced or eliminated before other tones (except another T3, where tone sandhi occurs), increasing its similarity to T4, and its perceptual confusability [5, 7]. Specifically, we present auditory stimuli consisting of monosyllabic words, drawn from a large corpus with connected speech, to listeners whose task it was to select one of four characters, corresponding to the word with its correct tone. The perceptual patterns for T3 and T4 are compared with those of T1 and T2, which are expected not to exhibit much confusability. Since our initial results revealed a potential bias favoring T4, we also present follow-up d' analyses that abstract away from the bias. Finally, we briefly discuss the acoustic properties of the corpus from which our stimuli are extracted, to further assess the relationship between the perception and production of the tones.

2. PRESENT STUDY

In connected speech, Mandarin Tones 3 and 4 may exhibit fairly similar (falling) contours, in particular when the final rise of T3 is reduced or absent. We may thus expect words with these tones to be relatively easy to confuse. By contrast, we expect other tone pairs, with more distinct contours, to be less subject to confusion. We thus test the following hypotheses:

(1) Hypothesis 1: Tone 3 and Tone 4 exhibit perceptual errors, where each tone is perceived as the other tone.

(2) Hypothesis 2: The rate of confusion between T3 and T4 is greater than that of the other tone pairs.

Although the focus here is on the perception of T3 vs. T4, to understand the characteristics of these tones that may lead to their confusion, we also briefly consider the acoustic properties of the stimuli, as independently determined for the corpus from which they were drawn [1]. Specifically, we discuss pitch (F0) and the phonation property HNR (Harmonic-to-Noise Ratio), the two measurements that found in [1] to be the main distinguishing properties for Mandarin Tones 3 and 4.

3. METHODOLOGY

3.1. Procedure

Seventeen native Mandarin speakers (11 females, ages 18-34), participated in a four-way forced choice perception experiment presented on a computer with e-prime software. The participants heard two repetitions of each target monosyllabic (CV) word, and selected one of four characters corresponding to the same syllable but with different tones. For example, for the auditory stimulus /ma/ with T3, the participants had to choose one of the following: T1 妈 'mother', T2 麻 'hemp', T3 马 'horse', T4 骂 'scold'. They indicated their responses on a response box with keys matching the positions of the characters on the screen, randomly displayed. Each participant heard half of the full set of stimuli (cf. Section 2.3), so the experimental session took approximately one hour.

3.2. Stimuli

The stimuli used in the perception experiment were monosyllabic CV words extracted from a large corpus originally collected for acoustic analysis [1]. This corpus, recorded in Beijing, consists of recordings of ten university-educated Mandarin speakers (4 females; mean age 22 years) producing real three-syllable (compound) words embedded in short dialogues, priming a reading with focus either on the target word or not (i.e., on a following word), for a total of 432 targets per speaker. The CV targets were flanked by syllables selected to minimize potential effects of tonal coarticulation; sequences of T3 were excluded to avoid tone sandhi.

For the present perception study, 36 target CV words, 12 of each of the vowels /i, u, a/ bearing T3 (e.g., 马 /mǎ/ 'horse'), and 36 similar CV words bearing T4 (e.g., 骂 /mà/ 'scold') were extracted from 8 of the speakers in the corpus (4 females). In addition, 36 similar CV distractors were extracted: 18 with T1 (e.g., 妈 /mā/ 'mother') and 18 with T2 (e.g., 麻 /má/ 'hemp'). The words were drawn equally from the three syllable positions of the tri-syllabic words and from the two dialogues (i.e., with and without focus) in the production corpus.

The items extracted from each voice (speaker) in the corpus constituted a separate block (N = 8) in the perception study. Each participant heard half of the blocks (2 female and 2 male voices), randomly distributed. Since pitch is a relative property, before beginning each block, the participants listened to two dialogues produced by the speaker, with items not used in actual experiment.

3.3. Analyses

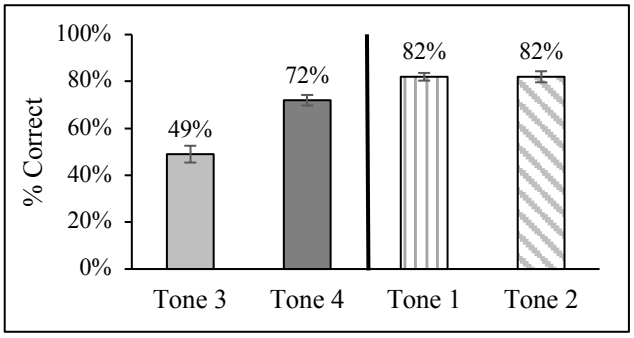
We first analyzed the responses for the rate of correct selection of the character corresponding to the auditory stimulus, and the distribution of errors. Since it appeared that there was a bias in the responses, we also conducted d' analyses for sensitivity. The acoustic analysis of the stimuli provided below is from the original production study.

4. RESULTS

4.1. Perception of Tones 3 and 4: correct responses

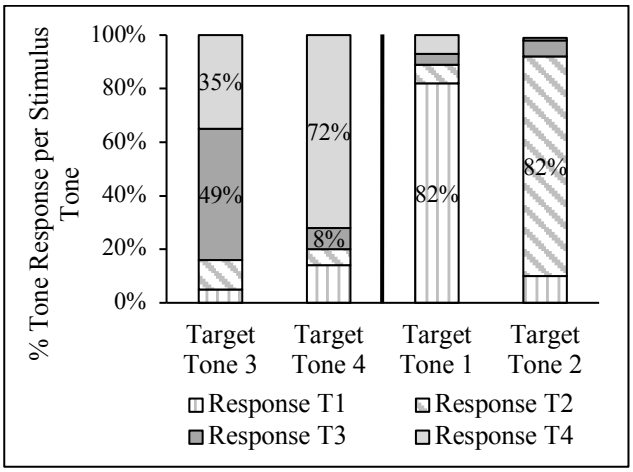
To address the initial question of whether Mandarin speakers consistently perceive Tones 3 and 4 when they are produced in connected speech, we first determined the percentage of correct responses. For comparison, we also examined the responses to the distractors, T1 and T2, which were expected to exhibit minimal perceptual confusion, although, as noted, they appeared in half as many stimuli as T3 and T4. The results for all four tones are shown in Figure 1. A generalized linear model analysis suggested a main effect of Tone on the rate of correct responses ($p < .001$). Pairwise t-tests with Bonferroni correction revealed that the rate for T4 is significantly higher than for T3 ($p < .001$). Comparison with T1 and T2 shows that both have higher correct rates than T3 and T4 ($p < .01$). As the figure suggests, T4's rate is closer to that of T1 and T2, than T3. From this analysis, it appears that while both T3 and T4 exhibit perceptual confusion, perceiving T3 in connected speech, without context, is especially difficult.

Figure 1. Overall correct responses for each tone.



To gain further insight into the response patterns, we also examined the distribution of incorrect responses. As Figure 2 shows, there is a bias in favor of T4 over T3. That is, when T3 is not perceived correctly, the vast majority of errors (35%) involve selection of T4. By contrast, T4 is only mistakenly perceived as T3 8% of the time. The responses for T1 and T2 are again provided for a baseline comparison, and again we see that T4 appears to be more similar to these tones, with no clear preference for another tone selection, than to T3.

Figure 2. Distribution of Responses for each Stimulus Tone.

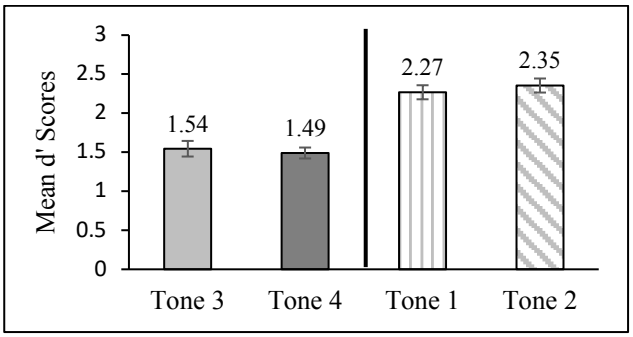


The fact that T4 is selected correctly 72% of the time, while T3 is also (incorrectly) identified as T4 35% of the time, however, suggests that there is a general preference for T4, and avoidance of T3 selection.

4.2. Sensitivity to Tones 3 and 4: d' Scores

Given the apparent presence of a bias in the data, we tested the participants' sensitivity to the tones using d' analyses, which take the bias into consideration. Figure 3 presents the d' values for T3 and T4, as well as those for T1 and T2 for comparison.

Figure 3. Mean d' Scores for each tone.



Pairwise t-tests between all pairs of tones with Bonferroni correction show that the sensitivity rates for T3 and T4 are not statistically different, nor are the rates for T1 and T2. The sensitivity rates for T3 and T4, however, are both significantly different from the rates for both T1 and T2 ($p < .001$).

This pattern is fundamentally different from that seen previously, where we considered the rate of correct responses, as well as the distribution of perceptual errors. That is, since the T4 bias was not addressed in those analyses, and T4 appeared to be more clearly perceived than T3, almost as well as T1 and T2. Only T3 exhibited considerable perceptual confusion. When the bias is incorporated into the analysis of the responses, however, the sensitivity rates for T3 and T4 are equivalent, and appear much lower than the sensitivity rates for T1 and T2.

5. DISCUSSION

5.1. Perceptual Confusion of Tone 3 and Tone 4

Returning to our hypotheses, we see that to some extent our determination of their confirmation rests on whether we are considering the accuracy of T3 and T4 selection, or sensitivity to the tones, as determined by a d' analysis.

In terms of accuracy, we find partial support for *Hypothesis 1*. That is, when considering the correct response rate, we do find perceptual errors between T3 and T4, but they are not reciprocal. Instead, they are primarily due to T3 being incorrectly perceived as T4 (i.e., 35%, compared to 49% correct T3 perception); T4 is incorrectly perceived as T3 only 8% of the time.

In general terms, *Hypothesis 2* is confirmed when the accuracy rates of T3 and T4 perception are compared with those of the distractors, T1 and T2, considered here as a baseline for what the perceptual pattern looks like when confusion is not expected. That is, overall, the confusion between T3 and T4 is greater than the confusion between T1 and T2. Closer examination of the errors, however, reveals that they are not evenly distributed. T1 and T2 both show minimal errors in which one tone is selected in place of the other (intended) tone: T2 incorrectly selected for T1 = 7%; T1 incorrectly selected for T2 = 10%. Similarly, as noted, T3 was incorrectly selected for T4 8% of the time, but T4 was incorrectly selected for T3 35% of the time.

Given the appearance of a bias, with T3 frequently being perceived as T4, but not vice versa, it was necessary to conduct a follow-up analysis using d', which takes into consideration correct and incorrect "hits" and "misses" in selection of each of the tones. When we evaluate our hypotheses in relation the d' sensitivity results, we arrive at a somewhat different view of the tone perception patterns. With regard to *Hypothesis 1*, we again see that there is considerable confusion associated with T3 and T4, but this time it is at essentially the same rate. That is, when the bias for T4 selection is excluded, the listeners' sensitivity to both tones is quite weak (around 1.5). By contrast, the sensitivity for both T1 and T2 is similarly strong (around 2.3). On this view, Hypothesis 2 is now confirmed, since we see considerable confusion between T3 and T4, but minimal confusion between T1 and T2.

5.2. Relation to Acoustic Properties

Since perceptual confusion indicates that the items in question, in this case, T3 and T4, are not clearly

distinguished in their production, the question that arises is what the acoustic properties of the stimuli are that might be responsible for their confusability. Since the stimuli for the present study were drawn from a corpus collected for an investigation of the production of the Mandarin tones, we now consider our perception findings in relation to the basic acoustic patterns revealed in production study [1].

In the acoustic analysis of the tones in the Mandarin corpus, measurements were made for several F0 properties (F0 = mean, Δall = change beginning to end, Δbeg = change in first half, Δend = change in second half), as well as for duration, energy, vowel centralization, and several phonation properties (HNR = harmonic-to-noise ratio, CPP = cepstral peak prominence, H1-H2). Binary Logistic Regression Analyses (BLRAs) were first used to identify which acoustic properties contributed significantly to the overall (i.e., with all of the measurements included) classification, or distinction, between the different tone pairs. Follow-up BLRAs were conducted with each of the properties found to be significant in the overall classification, using this property as the sole classifier, to indicate how strong a cue it is (by itself) for the distinction between a given tone pair. Table 1 provides the overall classification rate for T3 vs. T4, as well as for the "baseline" comparison of the classification rate for T1 and T2. In addition, the classification rates are provided for the three properties that yielded the most successful classifications (> 70%) on their own.

Table 1. Binary Logistic Regression Analyses: pairs of tones.

Tones	Overall	Individual Properties
T3 vs T4	85%	F0:79.5%, HNR:73%, Enr:71.6%
T1 vs T2	92%	F0:89%, HNR:73%, Δend:72.8%

The overall T3 vs. T4 classification rate is somewhat weaker than that of T1 vs. T2 rate, although it is moderately strong. The strongest individual classifier for T3 vs. T4 is F0 (79.5%), but it is closely followed by a phonation property, HNR (73%). Neither is very strong on its own, while in the case of T1 vs. T2, F0 accounts for most of the distinction on its own (89%); the next property, HNR, only accounts for 73%. Additional detail regarding F0 and HNR, the two main properties in both tone contrasts, is provided by Figures 4 and 5.

Figure 4. Normalized F0 for each tone

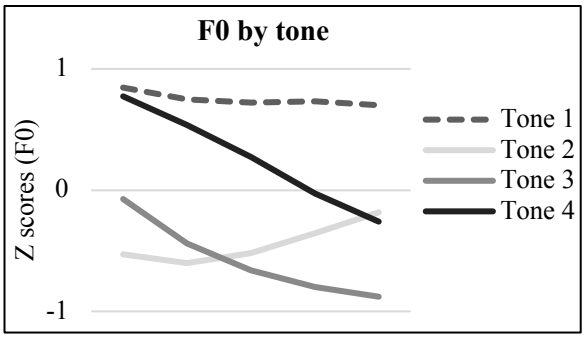
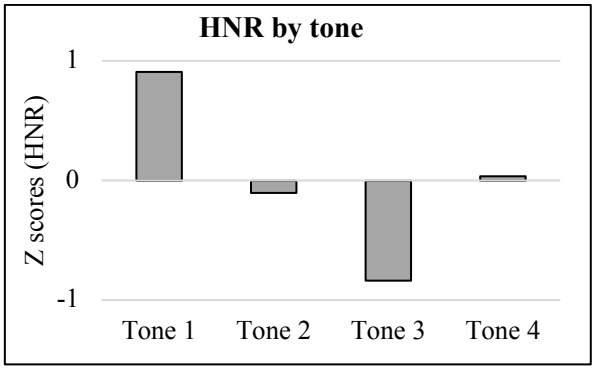


Figure 5. Normalized HNR for each tone



The patterns of the tones' acoustic properties show that even though F0 height and creaky phonation may serve as cues in the perception of T3 and T4, the similarity of their F0 contours nevertheless leads to confusability. This suggests that the F0 contour is a more important cue in the perception of tones in Mandarin than F0 height and phonation.

6. CONCLUSIONS

Our examination of the perception of Mandarin Tone 3 vs. Tone 4 in words drawn from connected speech shows considerable confusion, indicating that their acoustic properties are less distinct than their prescriptive descriptions, based primarily on words produced in isolation. When accuracy was considered, it appeared that T3 fared much better than T4; however, when a T4 bias was removed in a follow-up d' analysis, sensitivity to both T3 and T4 was quite low. By comparison, the T1 vs. T2 distinction was robust, even with our stimuli drawn from connected speech. Examination of the acoustic patterns observed in the corpus from which the perception stimuli were extracted [1], demonstrated, moreover, that the weaker T3 / T4 distinction may be largely attributed to their rather similar F0 contours. At the same time, however, their overall distinction may be somewhat enhanced by the CP found with T3, which combined with F0 (and Enr) in the BLRAs brought the distinction to 85%.

7. REFERENCES

[1] Athanasopoulou, A., Vogel, I. In preparation. The acoustic properties of Mandarin tones: effects of focus and syllable position.

[2] Boersma, P., Weenink, D. 2008. Praat. http://www.fon.hum.uva.nl/praat/

[3] Cao, R., Sarmah, P. 2007. *A perception study on the third tone in Mandarin Chinese*. Unpublished. University of Florida.

[4] Cao, R. 2012. *Perception of Mandarin Chinese Tone 2/Tone 3 and the role of creaky voice*. Unpublished PhD Dissertation. University of Florida.

[5] Chao, Y. R. 1965. *A grammar of spoken Chinese*. CA: University of California Press, 27.

[6] Gårding, E., Kratochvil, P., Svantesson, J. O., Zhang, J. 1986. Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech* 29(3), 281-293.

[7] Gårding, E. 1987. Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica* 44(1), 13-29.

[8] Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. *Phonetica* 20(2-4), 63-67.

[9] Moore, C. B., Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America* 102(3), 1864-1877.

[10] Shen, X. S., Lin, M. 1991. A perceptual study of Mandarin tones 2 and 3. *Language and speech* 34(2), 145-156.

[11] Shih, C., Kochanski, G. P. 2000. Chinese tone modeling with Stem-ML. In *Sixth International Conference on Spoken Language Processing*.

[12] Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *ICPhS XVII*, 1846-1849.

[13] Xu, Y. 1994. Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4), 2240-2253.

[14] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of phonetics* 25(1), 61-83.

[15] Yang, B. 2015. *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer Berlin Heidelberg.

[16] Yuan, J. 2003. Selective adaptation tests on Mandarin Tone2 and Tone3. In *Proceedings of ICPhS XV*,1719-1722.

[17] Yu, K. M. 2010. Laryngealization and features for Chinese tonal recognition. In *INTERSPEECH-2010*, 1529-1532.