PERCEPTION OF MANDARIN TONES BY CHINESE- AND ENGLISH-SPEAKING LISTENERS

XXX

XXX
XXX

ABSTRACT

This paper reports on two experiments that tested the hypothesis that native phonology may influence speech perception. Both experiments used natural speech tokens of Standard Mandarin tones and Chinese- and American English-speaking listeners. The results from both the AX discrimination and the degree of difference rating experiments show language-specific effects: the Chinese-speaking listeners' tone perception space was warped due to tone sandhi processes that neutralize two otherwise contrastive lexical tones. On the other hand, the English-speaking listeners showed phonetic listening, paying more attention to the similarity in pitch offset and onset between a pair of tones.

Keywords: speech perception, Mandarin tones, tone sandhi, neutralization, language-specificity.

1. INTRODUCTION

Standard Mandarin has four contrastive lexical tones: high level /55/, (mid) rising /35/, low (dipping) /214/ and high falling /51/, as described in Chao [1]. Henceforth, these tones will be labeled T55, T35, T214 and T51, respectively. Note, however, T214 usually does not reach level "4". In the stimuli used in our experiments, T214 is mostly a low tone, without much final rise. Chao [1] mentions a tone sandhi involving two consecutive T214s:

(1) T214 Sandhi: /T214.T214/ → [T35.T214]

Since an underlying /T35.T214/ sequence is also realized as [T35.T214], the paradigmatic contrast between T35 and T214 is lost before a following T214, resulting in surface pairs that are perceptually indistinguishable. Thus, /hao^{214}.mi^{214}/ 'good rice' and /hao^{35}.mi^{214}/ 'millimeter' are ambiguous when spoken in isolation, as both surface as [hao^{35}.mi^{214}].

In this paper we discuss results from two perceptual experiments designed to determine whether such a tone sandhi rule has an impact on tone perception by Chinese listeners.

2. Experiment 1: AX discrimination

2.1. Procedures

The participants were ten Chinese and thirteen American English (AE) listeners. All of them heard 140 pairs of naturally recorded stimulus tones carried by /bao/ in seven sections. The tones in a pair were either identical or different. These were played through headphones (at a 300ms inter-stimulus interval and a 2000ms inter-pair interval) at a comfortable listening level and with no background noise. The listeners had to decide whether they heard two identical or different tones by pressing the "same" or "different" button on a box. Both the judgment accuracy and reaction time (RT) were recorded as results.

2.2. Predictions

It was predicted that if T35 and T214 were more confusable than other pairs of tones, when listeners heard tone pairs involving T35 and T214, (i) they would make more mistakes, and (ii) they would take longer to make the judgment, on the assumption that the shorter the perceptual distance, the longer the RT, as shown in Shepard [2].

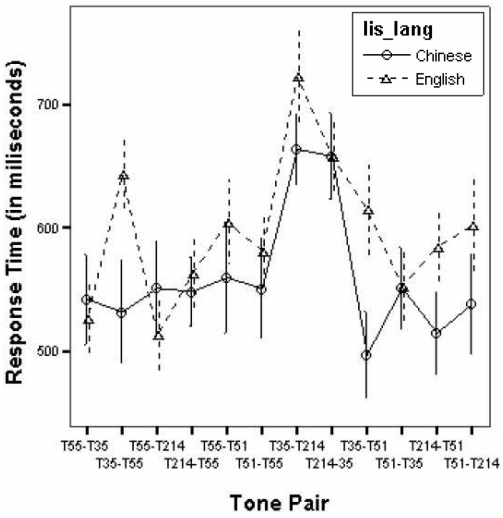
2.3. Results and analysis

The results from the AX discrimination experiment show that, for both the Chinese and AE listeners, T35 and T214 were indeed perceptually more confusable than any other tone pairs. In terms of the mistakes that listeners made, there was no statistically significant difference among the tone pairs, as error rates were fairly low for both listener groups, although the pairs involving T35 and T214 did attract more errors than the other tone pairs.

We then analyzed the RT data for the correct "different" responses, as we were interested in unveiling the perceptual distance between a pair of different tones.

Figure 1 shows the RT plots for the two listener groups. The repeated measures ANOVA (with all 12 non-identical tone pairs as the within-subject variable, and the two listener groups as the between-subject variable) did not find significant between-subject language effect ($F(1, 21) = .76$, $p = .393$). But pairs T35-T55 and T35-T51 turned out to be significantly different between the listener groups in the T test ($t = -2.136$, $p = .045$, and $\eta^2 = 0.178$; $t = -2.254$, $p = .035$, and $\eta^2 = 0.195$, respectively). The ANOVA found a significant effect with tone pair types, $F(7.487, 157.221) = 13.382$, $p < .001$, partial $\eta^2 = .389$. There was also a significant effect with the interaction of listener language and tone pair, $F(7.487, 157.221) = 3.295$, $p = .002$, partial $\eta^2 = .136$.

Figure 1: RT for the correct "different" responses. Error bars show one standard error.



Within-group pairwise comparison of the RT data showed that for the Chinese listeners, tone pairs T35-T214 and T214-T35 were the most confusable and were significantly different from all other pairs ($p < .05$), while pair T35-T51 was the least confusable and significantly different from all other pairs except for pairs T35-T55 and T214-T55. Although the AE listeners also found pairs T35-T214 and T214-T35 to be the most confusable, pair T214-T35 was not significantly different from T35-T55, T35-T51 and T51-T214. The AE listeners also found three tone pairs to be the least confusable and significantly different from most other pairs ($p < .05$), namely T55-T35, T55-T214 and T51-T35, which do not stand out in the Chinese listeners' data at all (see Figure 1).

A closer look at these patterns reveals that the AE listeners used the pitch onsets and offsets as phonetic cues to discriminate the tones: the more similar these points are for a pair of tones, the more confusable the pair was for them. Such was the case for T35-T55 and T35-T51. On the other hand, the Chinese listeners, with lexical tone categories, may have perceived the f0 contour on a monosyllable as an indivisible unit and may have ignored phonetic details to a certain extent, which explains why T35-T51, one of the more confusable pairs for the AE listeners, was the least confusable for them.

Interestingly, these strategies were not always to the advantage of either group of listeners: for T55-T35 and T55-T214, the AE listeners were better able to use the phonetic cues and responded faster. But the Chinese listeners were able to use pitch contour information more efficiently in pairs T35-T55 and T35-T51. This difference in processing strategies is a very telling one, because it suggests that the long RTs for T35-T214 and T214-T35 might have resulted from different factors for the two listener groups: if the Chinese listeners' perception was not affected by the phonetic similarity between the tones as was the AE listeners', the T214 sandhi in their native phonology could have played a role.

To uncover the factors affecting listeners' tone perception, we performed an individual differences (weighted Euclidean distance) multidimensional scaling (INDSCAL) analysis [3], on the correct "different" response RT data. Following Shepard [2], RTs were converted to perceptual distances using the reciprocal function:

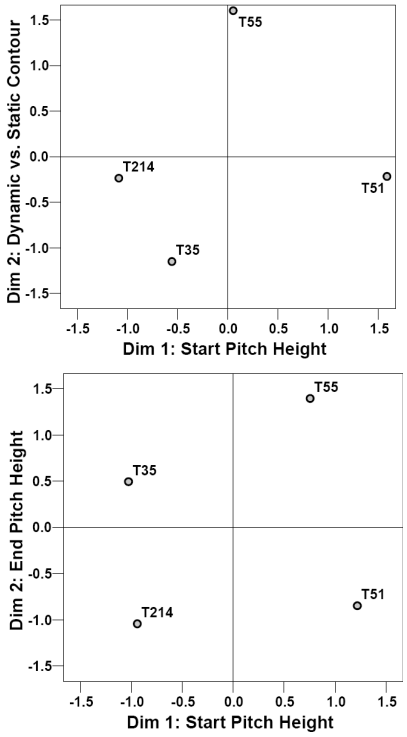
(2) Distance = 1/RT

and the resultant data were entered into one square (asymmetrical) distance matrix for each listener. Assuming the distance between a "same" pair to be zero, the data were analyzed at the measurement level "ratio" and compared "unconditionally" [4].

Figure 2 shows one perceptual space for each of the two listener groups, with fairly low stress levels and high percentages of variance accounted for ($RSQ_C = 0.89$ and $RSQ_{AE} = 0.91$). Dimension 1 in both spaces can be interpreted as "onset pitch height", as the tones at one end have higher starting pitch than the ones at the other end, although the Chinese listeners' space is slightly tilted. Dimension 2 in the AE listeners' space corresponds nicely to "offset pitch height", as the

tones at one end (i.e. T214 and T51) have lower pitch offsets than the ones at the other end (i.e. T35 and T55). This means that they perceived T35 and T214 as being similar for the same reason they did T35 and T55: the starting and/or ending pitch of the first syllable match the starting pitch value of the second syllable in a pair. In the Chinese listeners' space, there is an interesting "twist": T214 and T35 switched position along dimension 2. Since along this dimension T55 is separated from the three contour tones, it was interpreted as "dynamic vs. static contours". This shows that the Chinese listeners have indeed employed different processing strategies.

Figure 2: Two-dimensional perceptual tone spaces for the Chinese (upper panel; stress = 0.189, RSQ_C = 0.89) and AE (lower panel; stress = 0.169, RSQ_{AE} = 0.91) listeners.



In addition, the (relative) distance between T35 and T214 is shorter in the Chinese listeners' space. This is quite surprising because the Chinese listeners are more experienced with tone discrimination and should be expected to distinguish tone categories better than the AE listeners. Recall that in the within-group pairwise comparisons the Chinese listeners treated T35-T214 and T214-T35 as being significantly more confusable than all other tone pairs. This indicates that the category boundary between T214 and T35, i.e. the two tones involved in the T214 sandhi, may have been blurred by the neutralizing T214 sandhi rule, causing perceptual difficulty for the Chinese listeners that cannot be explained solely on grounds of phonetic similarity.

3. Experiment 2: degree of difference rating

In a simple AX discrimination task, confusability may reflect mainly auditory similarities among the tones (see, e.g. [5]). In Experiment 2, a degree of difference rating task was used in order to tap linguistic processing.

3.1. Procedures

Twenty-one (21) Chinese (Beijing) listeners and thirty (30) AE listeners participated in this experiment. They heard 192 pairs of naturally recorded tones carried by /ba/ in six sections. Otherwise, the design was similar to that in Experiment 1, except that the listeners rated the degree of difference between a pair on a "1" (very similar) to "5" (very different) scale subjectively.

3.2. Predictions

It was predicted that, if tone sandhi did not have an effect on native tone perception, Chinese and AE listeners' degree of difference rating would show similar patterns, both being influenced by phonetic similarity in pitch.

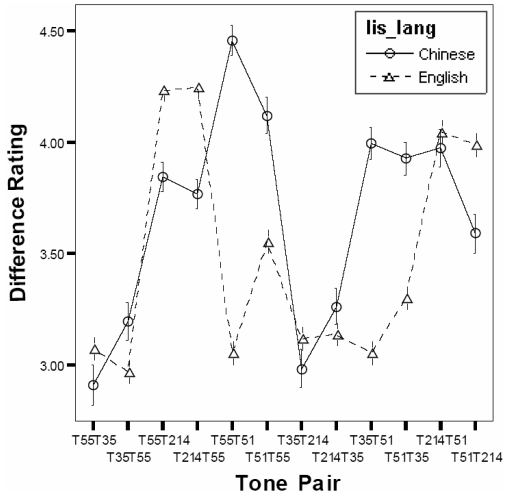
3.3. Results and analysis

The repeated measures ANOVA, with "tone pair type" as the within-subject variable (12 levels) and "listener group" as the between-subject variable (two levels), revealed a significant between-subject language effect, $F(1, 423) = 13.044$, $p < .001$, partial $\eta^2 = .03$. That is, how a listener rated the tonal difference was at least partially dependent on his/her native language. There was also a significant effect with tone pair types, $F(10.025, 4240.384) = 75.929$, $p < .001$, partial $\eta^2 = .152$. The interaction of language and tone pair types was significant as well, $F(10.025, 4240.384) = 40.609$, $p < .001$, partial $\eta^2 = .088$.

As can be seen in Figure 3, while there is similarity in the ratings by the two listener groups, there are significant differences (T test, $p<0.05$). When we examined tone pair effects within each listener group (ANOVA, pairwise comparison, $p<0.05$), the most significant overall pattern

became clear: for Chinese listeners, pairs T55-T35, T35-T55, T35-T214 and T214-T35 were the most similar, while for AE listeners, T55-T214, T214-T55, T214-T51 and T51-T214 were the most dissimilar. That is, the high tones were very different from the low tones for AE listeners.

Figure 3: Subjective degree-of-difference ratings by the Chinese and AE listeners. Error bars show one standard error.



While the AE listeners also rated T55-T35, T35-T55, T35-T214 and T214-T35 at the low end of the difference scale, they perceived *T51-T55*, *T55-T51*, *T35-T51*, and *T51-T35* to be very similar as well. Again, we notice similarities in the pitch offsets and onsets for all of these tone pairs.

The Chinese listeners, on the contrary, rated *T51-T55*, *T55-T51*, *T35-T51*, and *T51-T35* as "very different". Therefore, a different explanation is necessary for their low difference rating for T55-T35, T35-T55, T35-T214 and T214-T35. We may point to the T214 sandhi rule to account for the T35 and T214 pairs. But what about the T55 and T35 pairs? As it turns out, there exists another tone sandhi rule for Chinese speakers from Beijing:

(3) The T35 sandhi rule:
/T55.T35.Tx/ → [T55.T55.Tx], or
/T35.T35.Tx/ → [T35.T55.Tx],

where Tx is any non-neutral tone [1]. Since underlying /T55.T55.Tx/ and /T35.T55.Tx/ also surface with [T55] medially, the contrast between T35 and T55 is neutralized in this environment. In other words, the Chinese listeners rated any two tones as being quite different, except when a pair of tones is involved in a sandhi rule. (This rule did not affect the Chinese listeners in Experiment 1, as not all of them were from Beijing.)

4. CONCLUSION AND DISCUSSION

The results from the AX discrimination and degree of difference rating experiments reported here suggest that neutralization rules, such as the Mandarin T214 and T35 tone sandhis, may shorten the perceptual distance between two contrastive lexical tones in native Chinese perception, in contrast to the AE listeners' perception, which was influenced by phonetic similarity.

This finding poses interesting questions for models of speech perception and language acquisition. First, the fact that language-specific effects surfaced even in the AX discrimination task provides supporting evidence for the neural model proposed by Guenther and colleagues [6, 7], which hypothesizes that linguistic experience may lead to warping in the auditory cortex such that between-category perception is enhanced and within-category discriminability reduced. Since such an auditory cortical map has a neurophysiological basis, its language-specific "landscapes" should be reflected in low-level auditory processing, albeit not to the degree found in linguistic processing. Second, *ceteris paribus*, would acquisition of the tones or sounds involved in neutralization rules be hindered compared with other tones or sounds in the same system? Maybe only longitudinal studies can answer the question.

5. REFERENCES

[1] Chao, Y-R. 1965. *A Grammar of Spoken Chinese*. Berkeley & Los Angeles: University of California Press.

[2] Shepard, R. N. 1978. The circumplex and related topological manifolds in the study of perception. In Shye, S. (Ed.), *Theory construction and data analysis in the social sciences*. San Francisco: Jossey-Bass.

[3] Carroll, J. D., Chang J.-J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. In *Psychometrika*, Vol. 35, No. 3, 283 – 319.

[4] Takane, Y., Young F. W., DeLeeuw, J. 1977. Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.

[5] Pisoni, D., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. In *Perception & Psychophysics* 13, 253-260.

[6] Guenther, F. H., Gjaja, M. 1996. The Perceptual Magnet Effect as an Emergent Property of Neural Map Formation. *J. Acoust. Soc. Am.* 100, 1111-1121.

[7] Guenther, F. H., Bohland, J. W. 2002. Learning Sound Categories: A Neural Model and Supporting Experiments. In *Acoustical Science and Technology* 23(4), 213-220.