Language-specificity at early levels of Mandarin tone perception by Chinese listeners

Tsan Huang
University of Buffalo

Keith Johnson
UC Berkeley

1 Introduction

The influence of perception on phonology and historical sound change has been discussed by various researchers from very early on (e.g., Trubetzkoy, 1969 (1939)). In Jakobson, Fant, & Halle (1952), perceptual features are treated as primary (see also Jakobson & Halle, 1956). But the generative tradition of phonology since Chomsky & Halle (1968) centers on articulatory definitions of distinctive features (but cf. Liljencrants & Lindblom, 1972; Ohala, 1981). Now a revival of interest in the interplay of perception and phonology seems to be in progress. In research work done in the past two decades or so, Kohler (1990), Hura, Lindblom & Diehl (1992), and Steriade (2001) hold that phonological processes such as segmental reduction, deletion, and assimilation are perceptually tolerated articulatory simplification and that the direction of such processes is determined by perception. That is, these processes only take place when the output of such a change is found to be highly confusable with the input perceptually. In particular, Steriade (2001) proposes a universal perceptual map, which organises speech sounds in terms of their perceptual salience in different contexts. Johnson (2004) also presents data and analysis in support of such a perceptual map. If a certain contrast is perceptually weak in a certain position, synchronic phonology works to enhance or sacrifice it by way of epenthesis, metathesis, dissimilation, assimilation or deletion (Hume & Johnson, 2003). For example, vowel insertion between sibilants in some English plural noun forms (e.g., buses, bushes, judges; cf. cats, cans), metathesis of /skt/ to [kst] in Faroese and Lithuanian (Hume & Seo, 2004), and manner dissimilation of two consecutive obstruents in Greek (e.g., /kt/ → [xt] or /xθ/ → [xt]; Tserdanelis, 2001) serve to strengthen the syntagmatic contrast between these neighboring segments – although the latter two of these processes also result in paradigmatic contrast neutralization –, while n-lateralization in the /nl/ and /ln/ sequences in Korean and the optional /h/ deletion in Turkish sacrifice the perceptually weak contrasts for ease of articulation, leading to syntagmatic contrast neutralization in both cases (Seo, 2001; Mielke, 2003).

Ohala (1981) suggests that diachronic sound changes may occur due to the listener's misperception and reinterpretation of certain sounds or sound sequences. Instead of correcting distorted phonetic forms based on his knowledge of possible variants of the common underlying forms, the listener-turned-speaker may as well exaggerate the distortion, resulting in historical sound change (p.183). Janson (1983) further hypothesises a five-stage sound change process involving interaction between perception and production, using as an example the change of /r/ to /R/ in Norwegian (1983: 24). Guion (1998) suggested that the cross-linguistically common sound change of velar palatalisation is also perceptually conditioned. Examples of perceptual effects on sound change can also be found in historical tonogenesis and tone developments in the tone languages, where previously redundant pitch differences became contrastive when the

conditioning segmental contrast was lost (e.g. Maspero, 1912; Haudricourt, 1954a,b; Matisoff, 1973; Maran, 1973; Hombert, 1978; Hombert, Ohala & Ewan, 1979; Svantesson, 2001).

In this paper, we shall focus on the other aspect of the phonology-perception interplay, i.e. how phonology influences perception, and in particular how differences in tonal inventories and neutralization processes may affect tone perception by native listeners, in comparison with listeners who do not speak a tone language. It is important to test the influence of phonology on perception, i.e. language-specificity, because much of the work on the influence of perception on phonology assumes that the perceptibility of speech sounds is uniform across languages and that appeal to a universal perceptual space can be made.

There is evidence from past perceptual studies that listeners' perception of speech sounds depends on their linguistic experience. The inventories of contrastive sounds, the phonotactics of sound combination, and the phonological rules operating in the listeners' native languages all have an impact on speech perception. For instance, Japanese listeners, whose language has only one liquid sound, perceive the English /r-l/ distinction differently from American English (AE) speakers (see e.g. Miyawaki et al. 1975). Werker & Tees' (1984) study on the perception of Hindi voiceless unaspirated retroflex versus dental stops and Nthlakampx[1] velar versus uvular ejectives by native English speakers showed that listeners were unable to distinguish these foreign sounds contained in the /ʈa, ta/ and /k'i, q'i/ syllables. Hume, Johnson, Seo, Tserdanelis & Winters (1999) found that while consonant-vowel transitions seem to provide more place information for consonant place identification than stop burst for both AE and Korean listeners, the difference between the two kinds of stimuli is greater for Korean listeners. Hume *et al.* suggest that this is related to differences in phonological contrasts in these two languages; that is, Korean listeners with a three-way (tense, lax and aspirated) stop consonant contrast, which is cued in part by the duration of aspiration, seem to pay more attention to the CV transition between the burst and the vowel onset than do AE listeners, who have a two-way (unaspirated versus aspirated) stop contrast.

Hume & Johnson (2003) further suggest that four degrees of contrast should be taken into account in predicting the influence of native phonology on perception, namely fully contrastive, partially contrastive, allophonic, and non-occurring. They state that the influence of contrast on the perception of native sounds "is *not* all or none". In particular, a neutralization rule will shorten the perceptual difference between two categories that are otherwise contrastive in a language. Experimental data provide supporting evidence for such a statement. Fox (1992) found that English listeners fare poorly in identifying or discriminating vowels in the context of /hVr(d)/. Fox suggests that knowledge of the phonological rule that neutralises vowel contrast in this context may have affected the ability of listeners to make perceptual decisions about vowel quality. Harnsberger (2001) and Boomershine, Hall, Hume & Johnson (2008) present evidence for the claim that allophony can also affect perception.

The studies of segmental perception mentioned above indicate that speech perception is to a large degree language-specific, i.e. that phonology influences the perception of speech fairly extensively. Past studies have suggested that it is also the case in tone perception. Gandour (1983, 1984) and Lee, Vakoch & Wurm (1996) showed that for speakers of tone languages, differences in lexical tone inventory may play a role in tonal perception. In Gandour's (1983,

[1] Also referred to as Thomson, Nthlakampx is one of the Inner Salish languages spoken by the First Nations people of British Columbia.

1984) study using 19 synthesized f0 stimuli (five levels, four rising, four falling, three falling-rising, three rising-falling), 200 speakers/listeners of Mandarin (Taiwan, four tones), Cantonese (Hong Kong, six tones), Taiwanese (six tones), Thai (five tones), and English (no lexical tones) made dissimilarity judgments on tonal pairs. Results show that the tones were rated significantly differently by tone vs. nontone language speakers, by Thai vs. Chinese (Mandarin and Taiwanese) speakers, and by Cantonese vs. Mandarin and Taiwanese speakers. Such differences were attributed in part to differences in tonal inventories. In particular, tone height seems to be more important for English listeners, while Thai listeners attached most importance to the direction of f0 contour (i.e. rising versus falling). When the tone language groups were compared, Cantonese listeners seem to utilize mainly the dimension of tone levels/heights, which is not surprising, given that four of the six Cantonese tones have basically level contours. Lee et al. (1996) used naturally recorded stimuli of Cantonese and Mandarin tones on word and nonword syllables and had Cantonese, Mandarin (Taiwan and Mainland) and English (US) listeners participate in two "same"/"different" discrimination experiments. They found that Cantonese and Mandarin listeners were better at discriminating tones in their own dialect and that the tone language speaking listeners were better able to discriminate the tones than the English group.

Gandour (1981, 1983, 1984) suggests that tone sandhi rules may also influence tonal perception. Using Individual Difference Multidimensional Scaling (INDSCAL, Carroll & Chang, 1970), Gandour (1981) analyzed confusion data from native listener identification of naturally produced Cantonese tones. He found that the high falling tone was placed midway between the level and the contour tones in the perceptual tone space. He argued that this was due to the fact that this tone has a high level allotone in Cantonese. Although the allotone was not present in the stimuli, allophony still interfered with listeners' perception. The effect of the same allophonic alternation showed up in Gandour's (1983; 1984) experimental data, where Cantonese listeners perceived a /44/ (high level) contour to be similar to a /53/ (high falling) contour. In the same data, Mandarin listeners perceived the /44/ contour to be similar to /35/ (rising), which, as Gandour argued, was due to the existence of the allophonic rule that turns a rising tone to a high level in Mandarin (e.g. Chao, 1965, 1968; Cheng, 1973; see also §2).

In the present study, we tested the interplay of phonology and speech perception in the domain of lexical tone perception in Standard (Beijing) Mandarin Chinese. One key innovation of our research is that we examine the cognitive level of processing at which language-specific perception may take place. In one experimental paradigm we permitted listeners time to reflect on their answers in a similarity rating task, while in another paradigm listeners were required to respond very quickly and our measure of phonetic similarity was response time. This task difference, plus a speech/nonspeech manipulation using sinewave analogs of speech, allowed us to infer something about the level of perceptual processing that is affected by phonology. The main body of this paper is arranged as follows. Section 2 provides some linguistic background to the study. Section 3 discusses three experiments and compares differences and similarities between the data obtained from these experiments. Finally, in §4 implications of the experimental results from this study are discussed in relation to two speech perception models, namely Guenther & Bohland (2002; also Bauer et al., 1996) and Johnson (2004).

2 Tones and tone sandhis in Mandarin Chinese

Mandarin has four lexical tones. Chao (1965, 1968) describes them as high level /55/, mid-rising /35/, low falling-rising /214/, and high falling /51/. The numbers indicate the idealized pitch values of these tones on a five-level scale (Chao, 1930).[2] For notational purposes in this paper, we shall refer to them with the labels T55, T35, T214 and T51, respectively. There are also the so-called neutral-toned syllables in Mandarin, which are not specified for tone underlyingly. The surface tone value of a neutral-toned syllable is determined by the pitch offset of the preceding full tone (Cf. Chao, 1965, 1968).

Underlying full tones may be modified under the influence of their tonal environment in Mandarin (see, e.g. Chao, 1965, 1968; Kratochvil, 1968; Cheng, 1973; Chen, 2000; Duanmu, 2007). This phenomenon is known as tone sandhi. As described by Chao (1965, 1968), T214 of Mandarin becomes T35 when immediately followed by another T214:

(1) T214 Sandhi Rule: /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, creating many homophonous surface pairs. Thus, /hao^{214}.mi^{214}/ 'good rice' is indistinguishable from /hao^{35}.mi^{214}/ 'millimeter', as both surface as [hao^{35}.mi^{214}].

Traditional analyses state that other phonetic variants of T214 in normal stress positions include [21] and [214], the first of which appears before all full tones except T214 (where the T214 sandhi applies) and the second of which appears in prepausal positions, especially sentence-finally (Chao, 1965, 1968).

Chao (1965, 1968) also discusses a second tone sandhi rule, where T35 becomes T55 when following a T55 or T35 and preceding a full-toned syllable. Chao considers this rule to be "of minor importance" (1965: 35). Unlike the T214 sandhi, the T35 rule is not taught to second language learners.

(2) T35 Sandhi Rule

/T55.T35.Tx/ → [T55.T55.Tx], or
/T35.T35.Tx/ → [T35.T55.Tx],

where Tx is any non-neutral tone. According to Chao, the middle position of a trisyllabic phrase is relatively weak prosodically. As a result, the underlying low pitch onset of the sandhi-affected T35 gets deleted and the pitch contour is simplified to [55]. Note that the affected T35 does not have to be an underlying /T35/. For instance, a sequence of /T214.T214.T214/ may first be affected by the T214 sandhi rule (applied twice linearly from left to right), supposedly resulting in an inaudible middle stage of (T35.T35.T214), which is then further affected by the T35 sandhi rule, yielding the final surface sequence [T35.T55.T214]. Two familiar examples given by Chao (1965: 36) are: /cong^{55}.you^{35}.bing^{214}/ → [cong^{55}.you^{55}.bing^{214}] '(Chinese) onion pancakes', and /hao^{214}.ji^{214}. zhong^{214}/ → [hao^{35}.ji^{55}.zhong^{214}] 'quite a few kinds'. This sandhi also leads to paradigmatic neutralization: the contrast between T55 and T35 is lost here.

[2] Based on our small recorded database of ten (10) Beijing speakers, the tones are more likely /44, 24, 212, 51/ acoustically. The contours are nevertheless similar to Chao's description. Note that when the contour is fully realized, T214 may or may not have creaky voice in the middle.

The experiments in this study were designed to test the impact of such sandhi processes on perception by native Chinese listeners. American English (AE) listeners were included in the study as a control group to see (i) whether the tones involved in the sandhi rules share some property that makes them confusable for non-native listeners as well, (ii) whether the two groups of listeners perceive tones in a similar way, and (iii) whether the sandhi rules affect Chinese listeners' tonal perception. It is assumed that, if there is no effect of the listener's native phonology on perception, phonetic universality and human auditory capability should cause all listeners to behave the same.

Previous studies have shown that it is feasible to include non-tone language speakers in studies on tones. For example, Kiriloff (1969) found that, when asked to ignore the segmental element of the syllable and focus on the tones, non-native speakers' performance was quite good with an average of 87.5% correct identification. Although English is a non-tonal language, its stress-based prosodic system does utilize pitch as one way to distinguish stress accents, which may be realized as high, low, rising or falling contours and which can thus be very similar to the Mandarin tone contours (see e.g. Beckman, 1984). AE speakers may not know what to call the stimulus tones, but a paired comparison task does not require that listeners have labels for the items being compared. (Cf. Lee et al. (1996), where a small native speaker advantage was found.)

On the other hand, without lexical tone categories in their lexicon, AE listeners may actually enjoy an advantage; that is, they may be able to detect subtle pitch differences, which may be missed by Chinese listeners. Wang (1976) found that Mandarin-speaking listeners perceived synthesized stimuli along a level to rising contour continuum categorically, dismissing rises smaller than 9 Hz as negligible within-category variations. Similarly, Stagray & Downs (1993) reported that Mandarin-speaking listeners had significantly larger difference limens for frequency (DLFs) than English-speaking listeners around 125Hz, which approximates pitch utilized in a male voice. Their Mandarin listeners also had poorly-shaped psychometric functions close to chance response level, as opposed to regular ogive-shaped functions in AE listeners' data. Stagray & Downs concluded that Mandarin listeners had poorer differential sensitivity for frequency because frequency variations in the stimulus tones were perceived "as being within the same pitch range of a learned, level tone-phoneme category" (1993: 156).

3 The Experiments: Low Uncertainty AX Discrimination and Difference Rating tasks

This study tested the influence of tonal inventory and tone sandhi rules on native tone perception with three experiments involving different experimental methods and Chinese-speaking listeners (from the City of Beijing) as well as Midland American English-speaking listeners.

3.1. Experiment 1: AX discrimination task with sinewave tones

3.1.1. Participants and procedures

Experiment 1 employed a low uncertainty AX discrimination task. The purpose of this experiment was to try to tease apart the effects of raw acoustic similarity in the perceptual data from linguistic effects. Since in a low uncertainty AX discrimination task with non-speech tones one is likely to get only auditory perception, one may take the results from this task as the

perceptual baseline determined by raw acoustic similarity in the stimuli. Deviations from these results can then be seen as linguistic effects.

Eleven (six female, five male, average age 30) Chinese and thirteen (eight female, five male, average age 21) AE listeners participated in Experiment 1. All Chinese listeners were from Beijing. The AE listeners speak a Midland variety of American English. None of the listeners reported any history of speech or hearing problems. The Chinese listeners were paid a small amount of money for their participation, whereas AE listeners earned course credits.

All participants were tested in front of a computer in a quiet room, using the E-Prime program (Psychology Software Tools, Inc.). The stimuli were played through headphones. Experiment 1 involved an AX discrimination task and a short inter-stimulus interval (ISI) of 100ms. Such a short ISI tend to block high-level linguistic perception (e.g. Pisoni, 1973). Written and oral instructions were given to the two groups of listeners in their respective native language. In Experiment 1, after a listener responded correctly, a feedback message, detailing his/her reaction time (RT) and percentage correct, was displayed on the screen for 1500ms. Another 2000ms wait period followed. E-Prime then moved on to play the next pair of sounds. Both RT and response accuracy were recorded.

The stimuli were simple sinewave simulations of the natural speech tones recorded as monosyllables by a male Beijing Mandarin speaker in his early thirties. The stimuli were generated with a sinewave generator adapted from the C-code generously shared by Alex Francis and Howard Nusbaum at the University of Chicago. Specifically, the frequency of a single time-varying sinusoidal wave was modeled on the f0 of each of four recorded monosyllabic words /ba^{55}, ba^{35}, ba^{214}, ba^{51}/ (see Figure 1). The amplitude of the sinusoid was also modeled on the amplitude contour of these naturally produced syllables. The overall impression of the synthetic sinusoidal stimuli was that they were like low-pass filtered speech, but with the pure tone quality of a sinewave. We shall refer to these non-speech stimuli as T55, T35, T214, and T51.

A limited set of stimuli was presented in each trial block, with each block testing the discrimination of only two tones (e.g. T55 and T35 might be tested in block 1, T55 and T214 in block 2, and so on). Each of the four possible combinations of the two tones tested was repeated twice within each of the five (5) cycles in that block (e.g. block 1 might have four pairs, T55-T55, T55-T35, T35-T55, and T35-T35, all of which were repeated ten times, yielding 4 × 2 repetitions/cycle × 5 cycles = 40 pairs). There were 40 × 6 blocks = 240 trials in total in Experiment 1. The order of the blocks was randomized for different participants. There was a brief practice session, which contained just four pairs of tones.

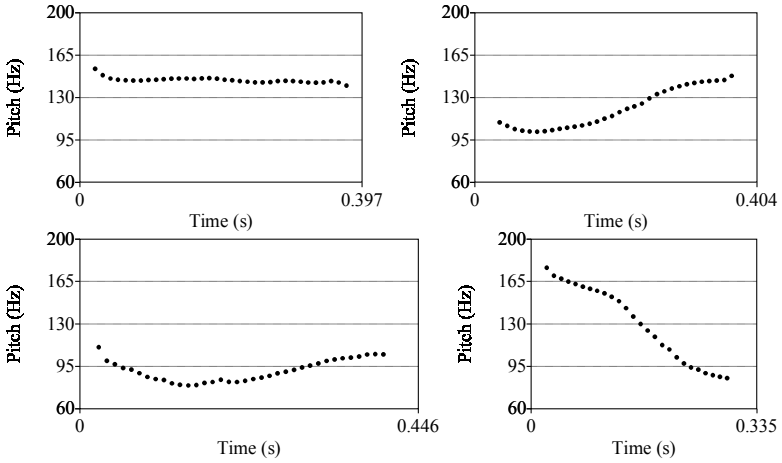


Figure 1 F0 traces of tones T55 (upper left panel), T35 (upper right), T214 (lower left), and T51 (lower right) as produced by a male Beijing speaker. The segmental makeup used in these recordings is /ba/. Lengths of the X-axes in these panels reflect approximately the relative lengths of the tones. These natural speech tokens were used in Experiments 2 and 3. They also served as templates for the synthetic sinewave stimulus tones used in Experiment 1. Note that the T214 tonal contour is fully realized, as it was produced in a prepausal position, although the final rise is still not to the level of "4" as the traditional analysis and label indicate, rendering T214 a rather low tone.

3.1.2 Results and Analyses

Error rates were low for both listener groups in Experiment 1. The Chinese listeners' overall error rate was 5%, with pairs T51-T55 (9%) and T35-T214 (13%) drawing the most errors. But these larger numbers were due to a very high error rate of one listener in each case (50% and 80%, respectively). When these outliers are thrown out, the error rates for these pairs conform to the overall error rate. Similarly for AE listeners, the overall error rate was 7%. The most errors were made with pairs T35-T55 (12%), T55-T51 (11%) and T51-T55 (9%). But again, these were attributable to one or two listeners' high error rate in each case. (Since the RTs were not outliers in these cases, these participants' data were included in the RT analyses.)

For the RT data, only results for correct "different" responses were analyzed using the repeated measures analysis of variance (ANOVA, general linear model), with "tone pair type" as the within-subject variable (12 levels) and "listener language" as the between-subject variable. For each tone pair, median RT values were computed for all individual participants. As can be seen in Figure 2 below, the overall RTs for the two groups are similar. There was a significant effect of the within-subject factor of tone pair type, [$F(8.6, 189.1) = 12.99$, $p < .001$, partial $\eta^2 = .37$]. No significant between-subject main effect of language group was detected by ANOVA in the RT data, [$F(1, 22) = .12$, $p = .73$, partial $\eta^2 = .005$]. But the tone pair by language interaction had a marginal effect, [$F(8.6, 189.1) = 1.95$, $p = .05$, partial $\eta^2 = .08$], suggesting that at least for some tone pairs there was a language effect. (Since Mauchly's Test of Sphericity was significant, $p = .001$ and epsilon $> .75$, the Huynh-Feldt correction is used here.)

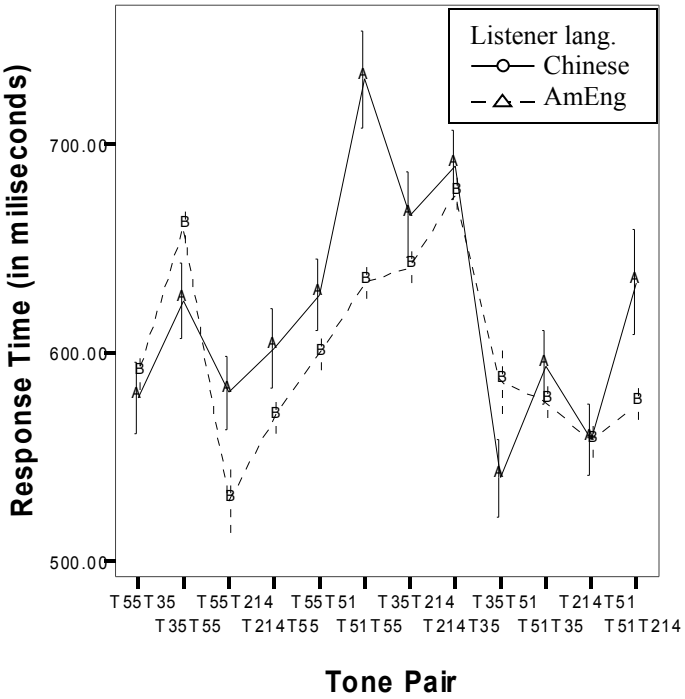


Figure 2 Response time plot from the ANOVA analysis for the experiment of AX limited stimulus set discrimination of sinewave tones for Chinese and AE listeners. (Error bars show one standard error.) Significant or marginal effects were found between the two language groups in the planned comparisons using an Independent Samples T test for T51-T55 ($p < .001$), T55-T214 ($p = .032$), T51-T214 ($p = .049$) and T35-T51 ($p = .069$). (A plot with data from pairs involving the same two tones collapsed is also presented in Appendix A.)

An Independent Samples T test was performed to further probe the possibility of a language effect as suggested by the marginal tone pair by language interaction in the ANOVA analysis. Significant or marginal effects were found between the two language groups for pairs T51-T55 ($p < .001$), T55-T214 ($p = .032$), T51-T214 ($p = .049$) and T35-T51 ($p = .069$). In the first three cases, AE listeners had shorter RTs, whereas in the fourth they had a slightly longer RT. These patterns can be accounted for if we assume that one strategy was used by AE listeners to compare the f0 offset of the first tone with the f0 onset of the second tone but that it was not adopted by Chinese listeners.

Within-subject pairwise comparison (ANOVA) for AE listeners' RT data showed that pair T55-T214 (with the shortest RT) and pair T214-T35 (with the longest RT) both differed significantly from four other pairs ($p < .05$). The rest of the pairs fell in the middle, with T55-T35, T55-T214, T35-T51, and T214-T51 showing no significant difference from any other tone pair. In Chinese listeners' data, the pairs are even less well separated due to large variances. We can only roughly derive three groupings: (i) T35/T214 and T55/T51 are the most confusable (i.e. with the longest RTs; see Shepard, 1978); (ii) T55-T35, T55-T214, T35-T51 and T214-T51 are the least confusable (i.e. with the shortest RTs; see Shepard, 1978); and (iii) T35-T55, T214-T55, T51-T35, and T51-T214 are not significantly different from any other pair. These patterns are summarized in Table I below.

Table I. Confusability ranking of tone pairs

	AE	Chinese
most confusable	T214-T35	T35-T214, T214-T35, T55-T51, T51-T55
	⋮	T35-T55, T214-T55, T51-T35, T51-T214
least confusable	T55-T214	T55-T35, T55-T214, T35-T51, T214-T51

Note that pairs T55/T51 are among the most confusable for Chinese listeners. Upon re-examining the sinewave stimulus tones, it was noticed that the falling portion of T51 was somewhat delayed as compared with its original speech template when the pitch traces were aligned relative to vowel onset and the durations normalized. In addition, the duration and intensity curves also differ between the sinewave tone T51 and its natural speech template: the sinewave tone was about 1/7 (or 40ms) shorter with a flat intensity line, while the intensity decreases sharply in the last 1/3 in the natural speech monosyllable (total duration = 340ms). Such discrepancies in the falling pitch contour and intensity envelopes may have contributed to the perceived similarity between T55 and T51. This oddity in the stimulus tone may have also resulted in longer RT for the AE listeners (comparing the results of experiment 1 with those of experiment 2), but the Chinese listeners seem to have been more influenced by the prolonged steady high tone of the 51 stimuli than were the AE listeners. If perception in this task was purely based on auditory discriminability, then one would not expect naturalness to have any impact. In addition, as mentioned above, AE listeners were faster at discriminating T55-T214 and T51-T214, two pairs that involve a high vs. low contrast, while Chinese listeners were slightly faster with T35-T51, a pair that involves two tones with a rising vs. falling contour.

3.2 Experiment 2: AX Discrimination of Natural Speech Tones

Experiment 2 was run with the same procedure (AX discrimination, ISI = 100ms, limited stimulus set within each test block) and the same two groups of AE and Chinese listeners. In addition, there were five new Chinese (Beijing) listeners (one male, five female, average age 27). The participants heard natural speech monosyllables this time. It was run within the same one-hour session right after Experiment 1 – except for five the Chinese listeners who only participated in Experiment 2 –, with a short break between experiments when the listener needed it. It was hypothesized that with more linguistic information in the natural speech stimuli, this experiment would result in more linguistic processing in Chinese listeners.

As in the Experiment 1 analysis, median RT values were determined for each participant for the repeated measures ANOVA. Language group profile RT plots are shown in Figure 3. No significant language effect was found, [$F(1, 27) = 2.5$, $p = .127$, partial $\eta^2 = .084$]. The within-subject tone pair type effect was significant, [$F(8.9, 239.3) = 6.96$, $p < .001$, partial $\eta^2 = .21$]. There was no significant "tone pair" by "language" interaction effect, [$F(8.9, 239.3) = 1.37$, $p =$

.2, partial η^2 = .05]. (Since Mauchly's Test of Sphericity was significant, $p < .001$ and epsilon > .75, the Huynh-Feldt correction is used here.)

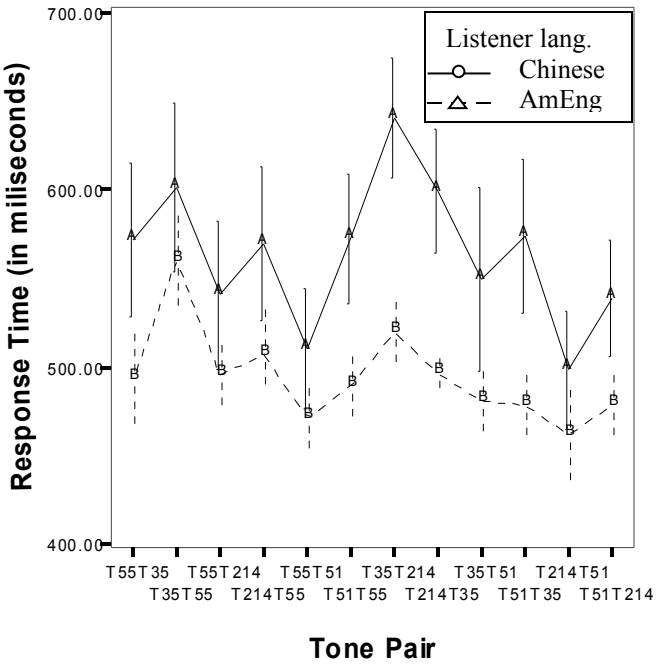


Figure 3 RTs (in milliseconds) for the correct "different" responses from the AX limited set discrimination task using natural speech stimuli. No significant language effect was found in the repeated measures ANOVA. (Error bars show one standard error.) But T test revealed significant differences between the two listener groups for pairs T35-T214 and T214-T35. (A plot with data from pairs involving the same two tones collapsed is also presented in Appendix B.)

With the low uncertainty design, error rates were again low, with the most errors occurring in T35-T214 (3.75%) for Chinese listeners and in T51-T55 and T35-T214 (6.92%) for AE listeners. While the RT functions for the two groups are largely parallel, the maximum at T35-T214 in the Chinese listeners' data is noticeably more prominent. When a planned between-group comparisons using the independent samples T test was performed on the RT data of all tone pairs, the result showed that the RTs for pairs T35-T214 and T214-T35 in Chinese listeners' data were significantly longer than those in AE listeners' data ($p = .009$ and $p = .015$, respectively), with over 20% of the variances accounted for in each case ($\eta^2 = .23$ and $\eta^2 = .3$, respectively). This indicates an influence from the T214 sandhi: Had the tone perception of both listener groups been affected by phonetic similarity only, Chinese listeners' RTs for pairs T35-T214 and T214-T35 should have conformed to the overall RT, that is, slightly longer than AE listeners' just as in the other tone pairs, but not significantly longer. After all, Chinese listeners were more experienced with distinguishing T35 from T214 than AE listeners. Why would they have performed significantly worse in this particular pair of tones than AE listeners, had phonetic similarity been the only factor affecting tone perception here?

Within-subject Pairwise comparisons (ANOVA on median RTs) on Chinese listeners' RT data revealed pairs T35-T214, T35-T55 and T214-T35 as the most confusable, with T35-T214 significantly different ($p < .05$) from six (6) other pairs, T35-T55 from four (4) pairs and T214-

T35 from three (3) pairs. On the other hand, pairs T55-T51 and T214-T51 were the least confusable, with each pair significantly different from three (3) other pairs. The other pairs fell in the middle. This is summarized in Table II below.

Table II. Confusability ranking of tone pairs in Chinese listeners' data.

most confusable	T35-T214, T35-T55, T214-T35
	⋮
least confusable	T55-T51, T214-T51

For AE listeners, although pairs T35-T55 and T35-T214 have relatively longer RTs, no significant difference was found between any two pairs. This means that the significant within-subject main effect of "tone pair type" came from the Chinese listeners' data. Again, the fact that pairs T35-T214 and T214-T35 were significantly more confusable than some other tone pairs for the Chinese listeners points to an effect of the T214 sandhi: since tone pairs involving T35 and T214 were not significantly more similar than any other tone pairs for AE listeners, phonetic similarity cannot be evoked to account for the perceptual pattern found in the Chinese listeners' data here.

There are noticeable differences between the results from Experiments 1 and 2. First of all, the overall RT is shorter with the speech stimuli in Experiment 2 for both Chinese and AE listeners, which may be seen as a practice effect (Werker & Logan, 1985), as Experiment 2 was run right after Experiment 1. This effect was more pronounced in AE listeners' RT data (see Figures 2 and 3 for comparison). RTs for pairs T55/T51, which were long in Experiment 1, are now among the shortest. Confusability rankings also differ in the two experiments for both Chinese and AE listeners, with the Chinese group showing relatively longer RTs for pairs T35-T214, T35-T55 and T214-T35 and the AE listeners showing no significant difference in RT between any two pairs in Experiment 2. RTs for pairs T35-T214 and T214-T35 were also significantly different between the two listener groups. These indicate an effect of the T214 sandhi on the Chinese listeners' tone perception. The relatively long RT for T35-T55 may be an indication of the T35 sandhi effect (see (2) in §2 above).

3.3 Experiment 3: Degree of Difference Rating of Natural Speech Tones

Following Shepard (1978) in assuming a positive correlation between RT and confusability, we tried to derive confusability rankings of pairs of tones from the RT data in Experiments 1 and 2. These experiments used a simple AX "same"/"different" discrimination task generally assumed to tap auditory processing (e.g. Pisoni, 1973; Macmillan, 1987; Johnson, 2004). Experiment 3 further investigated the confusability of Mandarin Chinese tones with a degree-of-difference rating task assumed to tap linguistic processing. That is, it was predicted that the difference rating task would bring out a stronger sandhi effect in the native Chinese listeners' tone perception.

Twenty-one (13 female, 8 male, average age 27) Chinese and thirty (15 female, 15 male, average age 20) AE listeners took part in Experiment 3. All Chinese listeners were from Beijing. Only three of these 21 Chinese listeners in Experiment 3 also participated in Experiments 1 and 2. All Chinese listeners were paid a small amount of money for their participation, whereas AE listeners earned course credits.

Experiment 3 consisted of six (6) blocks, each of which contained 32 natural speech stimulus tone pairs. (These were the same monosyllables used in Experiment 2.) Both block order and stimulus pair order within each block were randomized for each listener. All participants heard $32 \times 6 = 192$ pairs of identical or different tones. The 32 pairs in each block included the 12×2 different pairs and 4×2 identical pairs (i.e. each of the 16 possible pairings of the four tones was repeated twice in any of the test blocks). Each identical pair contained two repetitions of the same .wav sound file. The listeners were asked to listen carefully for tonal differences and rate the degree of difference on a "1" to "5" scale subjectively. The scale was described for them in the format shown in Table III. They were especially encouraged to use the full scale, instead of just "1" and "5". They were also asked not to think too much when they rated the differences, as we were concerned that we might get all "1"s and "5"s if they contemplated for too long. The five keys on the response box were labeled "1" through "5". Unlike in Experiments 1 and 2, no feedback was provided in Experiment 3. The setup was otherwise the same as that used in the first two experiments.

TABLE III. Rating scale described for the listeners on the Instruction sheet.

very similar	moderately similar	somewhat different	Moderately different	very different
1	2	3	4	5

Despite the instructions for using the whole scale of "1" through "5", some listeners used only "1" and "5". These data were discarded. As a result, only twenty-six (27) AE and eighteen (17) Chinese listeners' data were analyzed. The repeated measures ANOVA yielded the group profile plot in Figure 4. Obviously, the two listener groups had very different views on how similar or different the tones were. The overall pattern is that for Chinese listeners, only pairs T55-T35, T35-T55, T35-T214 and T214-T35 were relatively similar, while for AE listeners only pairs T55-T214, T214-T55, T214-T51 and T51-T214 were very dissimilar. Note that the pairs that were rated "similar" by Chinese listeners all consisted of tones involved in a tone sandhi, while the pairs that were rated "dissimilar" by AE listeners all involve the high vs. low pitch contrast. The Chinese listeners' rating of pairs T35-214 and T214-T35 as most similar is certainly consistent with the findings in Experiment 2, where it took them significantly longer than AE listeners to tell tones T214 and T35 apart. The patterns with pairs T55-T35 and T35-T55 are somewhat different in these two experiments. The degree-of-difference rating task, being more complex than a simple "same" or "different" discrimination one and with less time constraint, may have brought out a stronger effect of the T35 rule that neutralizes the contrast between T35 and T55 paradigmatically. Due to very different (sometimes opposite) maxima and minima,

ANOVA found no significant between-subject language effect, [F (1, 42) =1.77, p = .19, partial η^2 = .03]. But there was a significant effect with the within-subject factor of tone pair types, [F(5.61, 235.6) = 33.3, p < .001, η^2 = .442]. The interaction of listener language by tone pair types was significant as well, [F(5.61, 235.6) = 18.2, p < .001, η^2 = .303]. (Since Mauchly's Test of Sphericity was significant, p < .001 and epsilon < .75, the Greenhouse-Geisser correction is used here.)

The differences seen between the two language groups (Figure 4) showed up clearly in planned comparisons of the rating data using independent samples T tests. As reported in Table IV below, the ratings for pairs T55-T214, T214-T55, T55-T51, T51-T55, T35-T51, T51-T35, T51-214 and T35-T55 are significantly different for the two groups of listeners (p < .05). The largest rating disparity lies with pairs T55-T51 (t = 14.51, p< .001, η^2 = .34) and T35-T51 (t = 9.49, p < .001, η^2 = .183). In both cases, AE listeners rated the pairs as being similar, further supporting our analysis that AE listeners pay more attention to the start/end pitch points of the tones than do Chinese listeners, who were able to use the tone contour information in these cases.

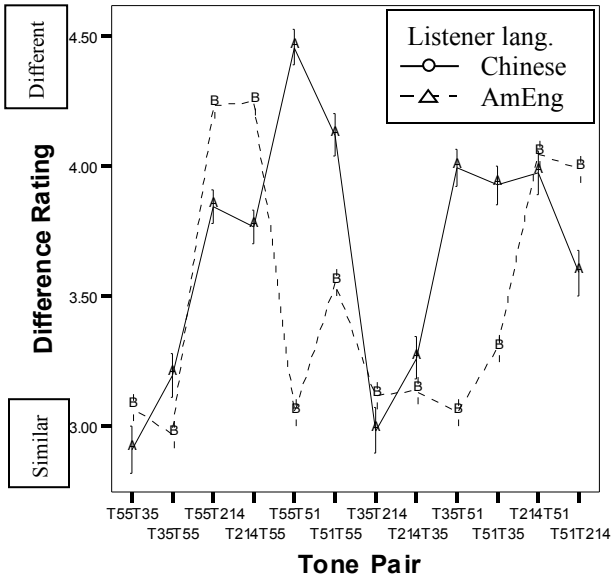


Figure 4 Subjective degree-of-difference ratings by Chinese and American English listeners. These group average values were computed from each listener's tone pair median values. "1" = "very similar", "5" = "very different". Error bars show one standard error.

TABLE IV Tone pairs for which AE and Chinese listeners' ratings were significantly different.

Tone pair	*t*	*p*	η^2
T35-T55	2.14	.033	.01
T55-T51	14.51	< .001	.34
T51-T55	5.39	< .001	.06
T35-T51	9.49	< .001	.18
T51-T35	6.45	< .001	.09
T55-T214	-4.51	< .001	.05
T214-T55	-5.45	< .001	.06
T51-T214	-3.73	< .001	.04

Within-subject pairwise comparison (ANOVA) revealed significant differences among tone pairs ($p < .05$). For the Chinese group, the most dissimilar pair was T55-T51 and the most similar pairs were T55-T35, T35-T55, T35-T214 and T214-T35 (see Figure 4). All other pairs fall in-between. Thus, the pattern is: Chinese listeners treated any two tones as being quite different, except when the two tones are involved in a sandhi rule, in which case the two tones were rated very similar.

For the AE group, at the same significance level, the most dissimilar pairs are T55-T214, T214-T55, T214-T55 and T214-T51, and the most similar pairs are T55-T35, T35-T55, T55-T51, T35-T214, T214-T35, T35-T51 and T51-T35. Pair T51-T55 falls in the middle of the scale but is not significantly different from pair T51-T35, so it may also be grouped with the more similar group. This is summarized in Table V below.

TABLE V Confusability ranking of tone pairs in the AE and Chinese listeners' degree-of-difference ratings data.

	AE	Chinese
most similar	T55-T35, T35-T51, T55-T51, T35-T214, T214-T35, T35-T51, T51-T35, T51-T55	T55-T35, T35-T55, T35-T214, T214-T35
		⋮
most different	T55-T214, T214-T55, T51-T214, T214-T51	T55-T51

As noted above, the common characteristic shared by the pairs rated as being more similar by AE listeners seems to be matching f0 onset and/or offset values, e.g. pairs T55-T51, T51-T55, T35-T51 and T51-T35. On the other hand, these same pairs were among the more dissimilar to Chinese listeners, as the contour shapes are very different for the tones involved in these tone

pairs. When the pitch onset and/or offset values are different, as in the case of pairs T55-T214, T214-T55, T214-T51 and T51-T214, AE listeners perceived the tones as being the most dissimilar, although these pairs do not stand out in Chinese listeners' data. Since order differences for most pairs of tones were negligible, we also show collapsed plot of the difference ratings in Appendix C, where it is even easier to see the general patterns: Chinese listeners had rated only the two pairs of tones involved in the T214 and T35 sandhis as "similar"; on the other hand, given time for reflection and (presumably) the influence of linguistic experience, AE listeners seemed to have used a "pitch height" strategy, and rated only the two pairs of tones involving a high vs. low contrast as "dissimilar".

3.4 General Discussion

The results from the AX degree-of-difference rating task (Experiment 3) show that Chinese listeners' tone perception was influenced by the tone sandhi rules in their native language: tone pairs T35-T214, T214-T35, T55-T35 and T35-T55 were rated as most similar because these tones are involved in the T214 or T35 sandhi (see §2), which leads to contextual neutralization of the tonal contrast between T35 and T214 or the contrast between T55 and T35. On the other hand, AE listeners rated pairs T55/T214 and T51/T214 as most dissimilar, showing that high vs. low was the most salient contrast in pitch for these non-tone language speakers.

The effect of tone sandhi on Chinese listeners' tone perception is quite remarkable in strength, because even a simple AX limited stimulus set discrimination using natural speech stimuli (Experiment 2) did not take it away completely: as revealed by the Independent samples T tests, it took the Chinese listeners significantly longer to make the "same"/"different" discrimination decision. But there was evidence that the Chinese listeners' attention was directed more toward the acoustic characteristics of the tones in such a simple task, as the maxima and minima on the two group RT plots align fairly well.

The two groups of listeners behaved even more alike in the limited stimulus set AX discrimination test using sine wave tones (Experiment 1), where no obvious sandhi effect was found. It is possible that, with the segmental makeup taken away and with just four stimulus tone tokens repeated over and over in Experiment 1, it was easy to focus attention on the acoustic properties of the stimuli. As a result, the data probably reflect mainly auditory perception. Although the Chinese listeners reported that they heard Chinese tones in this experiment using sinewave tones, it is rather doubtful that there was lexical activation involved in this task. Nevertheless, even in this experiment using synthetic stimulus tones, there were still some differences in how the two groups reacted to certain tone pairs, namely T55-T214, T51-T55, T51-T214 (shorter RTs for AE than for Chinese listeners), and T35-T51 (longer RTs for AE listeners). We also noted that the unnaturalness in the synthetic stimulus tone T51 affected Chinese listeners' perception but not AE listeners', which shows that perception in this task was not completely based on auditory discriminability; otherwise, one would not expect naturalness to have any impact here. As the aggregated RT plots (Figure 1) show, even in this simple task, the RTs for T35-T214 and T214-T35 are still somewhat longer relative to other tone pairs in the Chinese listeners' data than that in the AE listeners', albeit non-significant statistically.

As we noted above, when time was allowed for reflection in Experiment 3, AE listeners adopted a strategy, perhaps based on English intonation, to listen for pitch extremes and to

disregard contours. This enabled them to distinguish more easily (as reflected in a high score on the degree of difference rating scale) the two pairs of tones involving a high vs. low pitch contrast. Interestingly, it is less clear to what extend this strategic approach to the stimuli was used in the low uncertainty psychophysical tasks (Experiments 1 and 2), as AE and Chinese listeners looked more alike there – except for the sandhi pairs. This suggests that in a task similar to normal language use situations (Experiment 3) it may be impossible for a listener to not be influenced by his/her native language, even in cases where the stimuli are non-native.

4 Conclusion

As is evident from the experimental results reported above, linguistic experience can lead to language-specific patterns in speech perception. Different researchers try to account for such language-specific patterns in different ways. Steriade (2001) assumed a universal map of perceptual prominence that is not sensitive to linguistic factors. There is no doubt that general auditory capacities do not differ much among language learning children with normal hearing ability, no matter how diverse their linguistic backgrounds are. It is thus reasonable to assume that there exist universal patterns in perception of both speech and nonspeech auditory stimuli by speakers of different languages. However, our data suggest that the idea of a universal perceptual map may be an over-simplification.

It might be possible to account for much of the language-specific effects found in previous research on perception by noting that listeners of different languages have different phonological categories. In tasks that specifically ask listeners to use categorical knowledge (i.e. to tap long-term memory representations), it should be no surprise to find language-specific response patterns. This could be true even if the underlying sensory perceptual map is exactly the same for each listener, regardless of linguistic experience. For example, Carney et al. (1977) contend that categorical perception (e.g. Abramson & Lisker, 1970) may co-exist with general psychophysical perception under certain experimental conditions. Patterns not conforming to categorical perception in their discrimination data led Carney et al. (1977) to assert that a distinction should be made between attentional factors in perception and the perceptual capacities of human auditory organisms. Thus, language-specific effects such as those revealed in categorical perception may have resulted from different processing modes, namely a "general auditory" mode and a "phonetic" (phonological, in our terms) mode (see also Pisoni, 1973; Johnson, 1988). As a result, none of the data regarding the influence of phonology on perception negates the hypothesis that there is a universal perceptual map.

However, there exists another possibility that may be able to account for language-specific perceptual patterns, namely an auditory map as posited in the neural models advanced by Guenther and colleagues (Guenther & Gjaja, 1996; Guenther et al., 1999; Guenther & Bohland, 2002; Guentheret al., 2004) and Bauer, Der & Herrmann (1996). A central component in these neural models of perception is an auditory cortical map, whose formation is influenced by stimulus input and type of training. In particular, Guenther et al. (1999) found that categorical training in psychophysical experiments using nonspeech-like bandpass-filtered acoustic noise in different frequency ranges led to smaller cortical representation of (hence, decreased sensitivity to) stimuli in the training range, while discrimination training led to larger cortical representation (hence, increased sensitivity) in the training range. Functional magnet resonance imaging (fMRI) studies by Guenther & Bohland (2002) and Guenther et al. (2004) provided further supporting

evidence for this assertion. Greater temporal lobe activation was recorded when subjects heard non-prototypical examples of American English /i/ than when they heard prototypical examples of /i/ (Guenther & Bohland, 2002; Guenther et al., 2004). If we may further interpret Guenther and colleagues' results, their prototypical examples of /i/ can be seen as stimuli from the "categorization training range", except that the training was not done under laboratory conditions but was rather a listener's lifetime experience with his/her native language.

If an auditory warping similar to what Guenther et al.'s (1999) model describes existed, it would certainly serve the linguistic purpose well, as the warping directs neural activities to distinguishing between-category differences and to ignoring irrelevant within-category differences. And as pointed out by Guenther & Gjaja (1996), this would also enable a unified neural model for speech modalities and other sensory and motor modalities. Deutsch, Henthorn, Marvin & Xu (2006) report on an identification test on absolute pitch, i.e. 'the ability to name or produce a note of particular pitch in the absence of a reference note' (2006:116), among 88 Chinese-speaking and 115 English-speaking college musical students in Beijing and in the US, respectively. It was found that while it is the case that the smaller the onset age of musical training, the higher the probability of absolute pitch acquisition for both language groups, the Chinese students were significantly much more accurate in naming a pitch than their English-speaking counterparts in all subgroups of onset age. In other words, the Chinese musical students' linguistic experience with lexical tones may have helped them form musical tone categories. And structural MRI findings provide evidence for such a link (Schlaug, Jaencke, Huang & Steinmetz, 1995). Another recent study by Krishnan, Xu, Gandour & Cariani (2005) reported language-specific human frequency following response (FFR) patterns in a tone listening experiment by native speakers of Mandarin Chinese and English. The FFR data of the Chinese listeners reveal that these tone language speakers have a stronger pitch representation as well as a smoother pitch tracking than the English speakers. Krishnan et al. argue that such results support the hypothesis that there is neural plasticity even at the brainstem level. As a result, even during pre-attentive stages there may be language-specific pitch processing. Since language-specific effects were found in even the simplest of the three experiments reported here (i.e. Experiment 1 with a short 100ms ISI, limited set of stimuli testing within each block and a simple "same"/"different" discrimination task), our results provide yet another piece of empirical evidence for the hypothesis of an auditory cortical map: since the auditory cortical map has a neurophysiological basis and whose landmarks should be reflected in the perception data, regardless of the task.

Intuitively, one should still have the ability to perceive non-native contrasts and form novel categories in natural language situations; otherwise, adult acquisition of L2 would be more difficult than it is. Thus, auditory warping has to be reversible and the neural map re-arrangeable. Or perhaps a separate neural map is formed when the brain is trained with a separate set of stimuli, i.e. sounds from a foreign language. Recent research data on the plasticity of the human brain suggest that 'plasticity is a basic process that underlies neural and cognitive functioning' and that even in mature adult brain neural connectivity can still be 'altered by change in input to the system' (Stiles, 2000:241, 2000:266).

Under certain experimental conditions, some low memory demand tasks may not produce patterns conforming to patterns of the warped auditory map (e.g. Pisoni, 1973; Goldstone, 1998). Indeed, Guenther et al. (1999) did not find the perceptual magnet effect (Kuhl, 1991) using experimental procedures intended to promote a sensory-trace auditory processing mode, rather

than a context coding mode (e.g. Macmillan, 1987), especially a short ISI of 250ms and an AX discrimination task. Wang (1976) also found that extensive practice may shift categorical boundaries of tone-language speaking listeners closer to those of the non-tone language speaking listeners. In the present study, we found different degrees of language effect on tone perception. Guenther et al. (1999) offered no explicit account for these different task-dependent patterns. We can only infer from Guenther & Bohland's (2002) and Guenther et al.'s (2004) fMRI studies that these task-dependent patterns in experimental data not conforming to categorical perception or the perceptual magnet effect may be due to different degrees of activation in different auditory cortical areas in the temporal lobe and supratemporal plane and may be attributed to attentional factors.

On the other hand, Johnson's (2004) lexical distance model, although not explicitly denying the existence of auditory warping, tries to account for language-specific effects by referring to activation in the lexicon. In the lexical distance model, incoming signals are compared with phonetically detailed forms in the lexicon directly. Consequently, language-specific perceptual effects may simply emerge from the lexicon. The model computes overall perceptual distance (d) from two sources: 1) inherent auditory similarities between two stimuli (d_a), and 2) aggregated average difference in lexical activations by the two stimuli (d_l, computed as the difference in the amounts of activation of the lexicon caused by these stimuli, with a constant k gating the influence of this lexical distance on perception under different experimental conditions); i.e. $d = d_a + k \times d_l$. Because of the way the overall perceptual distance is computed, it is claimed that the model has the ability to distinguish discrimination performance from categorization performance. Discrimination performance can be found in a minimal uncertainty task of limited stimulus set or speeded AX discrimination with a short ISI such as Experiments 1 and 2 in our study (no lexical access assumed, perceptual distance computed almost exclusively from auditory distance). And categorization performance is found in tasks involving higher memory load such as AXB identification, the degree of difference rating as in Experiment 3 in our study (where it is hypothesized that lexical forms may be consulted for similarity judgments). Johnson's (2004) fricative perception data from a rating task as well as a speeded AX discrimination task by Dutch and AE listeners support this claim. The fact that the language effect is the most significant in the degree-of-difference rating task for both Chinese and AE listeners in our study provides further supporting evidence for the different degrees of lexical activation posited in Johnson's (2004) model.

Neither the neural model (Guenther & Bohland, 2002) nor the lexical distance model (Johnson 2004) explicitly discusses the issue of how neutralization rules may affect discrimination of two contrastive sounds (or tones) that are neutralized in certain phonetic environments. Within Guenther et al.'s neural model, we may imagine a "noisy" training condition under which stimuli categorized into an abstract representation of A may sometimes have to be categorized as A or B (e.g. [T35] to /T35/ or /T214/ because of the /T214.T214/ → [T35.T214] neutralization rule). As a result of this double-identity status of certain speech sounds, the contrast between the relevant tone (or phoneme) categories may be weakened and category boundaries less well defined. Within Johnson's lexical distance model, because of the cross-representation of two tones or sounds (e.g. /T35/ and /T214/ in our case), a [T35] or a [T214] input may activate lexical items containing either /T35/ or /T214/. Consequently, the difference in lexical activation, i.e. the lexical distance, between /T35/ and /T214/ is predicted to be smaller than if there is no such neutralization rule.

In summary, the findings of the three experiments reported here support the hypothesis that speech perception is language-specific, specifically that tone sandhi processes influence native listener tone perception at various levels. This is certainly consistent with the results reported in Gandour (1981; 1983; 1984), Deutsch et al. (2006) and Krishnan et al. (2005). The fact that language-specificity showed up in even the simplest task of speeded "same"/"different" discrimination with a short ISI in our study provide supporting evidence for the hypothesis that linguistic experience may lead to auditory warping (Guenther & Bohland, 2002): a neural map, once formed, should be reflected in perceptual patterns even before higher level linguistic processing is involved. However, different experimental tasks brought out different degrees of language-specificity (from the strongest tone sandhi effects found in our degree-of-difference rating Experiment 3, to a weaker effect found in the speeded discrimination tasks using natural speech and synthetic stimuli in Experiments 2 and 1), suggesting that speech perception may also be influenced by the degrees of lexical activation involved in different tasks (Johnson, 2004). This led us to the speculation that there may exist two types of linguistic effect: the first, found in tasks involving mainly auditory processing (as in our Experiments 1 and 2), comes from the auditory cortical map, where category boundaries between segments and lexical tones are defined by a particular segmental or tonal system (including allophonic or tone sandhi rules); the second, found in tasks involving higher level linguistic processing (as in our Experiment 3), comes from both the auditory cortical map and the more conscious consultation of phonological or tonal rules. In essence, this speculation is not very different from Johnson's (2004) lexical activation model, where the auditory and lexical effects interact to various degrees in different tasks. The only difference is that in Johnson's (2004) model, auditory distance is computed based on pure auditory similarity that is universal to speakers of any language, whereas in our view this purely universal auditory similarity may not exist, due to differences in the auditory cortical maps. As a result, language-specific effects may be found even in early levels of perceptual processing, through the category-defining influence of phonology during the formation of the auditory cortical map.

References

Abramson, A. & L. Lisker (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of phonetic Sciences* (Academia, Prague). 569-563.

Bauer, H.-U., R. Der & M. Hermann (1996). Controlling the magnification factor of self-organizing feature maps. *Neural Computation* 8. 757-771.

Beckman, M. E. (1984). Toward phonetic criteria for a typology of lexical accent. Ph.D. dissertation, Cornell University.

Boomershine, A., Hall, K.C., Hume, E. & K. Johnson (2008). The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher & K. Rice (eds.) *Contrast in Phonology: Perception and Acquisition*. New York: Mouton de Gruyter.

Carney, A. E., G. P. Widin & N. F. Viemeister (1977). Noncategorical perception of stop consonants differing in VOT. In *JASA* 62. 961-970.

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

Chao, Y.-R. (1930). A System of Tone Letters. *La Maître phonétique* 45:24-27.

Chao, Y.-R. (1965). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, M. Y. (2000). *Tone sandhi patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

Cheng, C.-C. (1973). *A synchronic phonology of Mandarin Chinese.* The Hague: Mouton.

Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York, Harper & Row.

Deutsch, D., Henthorn, T., Marvin, E., & Xu H.-S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *JASA* 119. 719-722.

Duanmu, S. (2007). *The phonology of standard Chinese*. 2nd edition. Oxford: Oxford University Press.

Fox, R. A. (1992). Perception of vowel quality in a phonologically neutralized context, in Y. Tokura, E. Vatikiotis-Bateson, & Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*, Tokyo: Ohmsha and IOS Press (jointly). 21-42.

Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. In *Journal of Chinese Linguistics* 9. 20-36.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. In *JPh* 11. 149-176.

Gandour, J. T. (1984). Tone dissimilarity judgments by Chinese listeners. In *Journal of Chinese Linguistics* 12:2. 235-61.

Goldstone, R. L. (1998). Perceptual learning. In *Annual Review of Psychology*, 49: 585-612.

Guenther, F. H. & M. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. In *JASA* 100. 1111-1121.

Guenther, F. H., R. T. Husain, M. A. Cohen & B. G. Shinn-Cunningham (1999). Effects of Categorical and Discrimination Training on Auditory Perceptual Space. In *Journal of the Acoustical Society of America*, 106. 2900-2912.

Guenther, F. H. & J. W. Bohland (2002). Learning sound categories: A neural model and supporting experiments. In *Acoustical Science and Technology* 23:4. 213-220.

Guenther, F. H., A. Nieto-Castanon, S.S. Ghosh & J. A. Tourville (2004). Representation of sound categories in auditory cortical maps. J*ournal of Speech, Language, and Hearing Research,* 47:1. 46-57.

Guion, S. G. (1998). The Role of Perception in the Sound Change of Velar Palatalisation. In *Phonetica* 55. 18-52.

Harnsberger, J. D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. In *JPh* 29. 303-327.

Haudricourt, A.-G. (1954a). De L'Origine des Tons en Vietnamien. In *Journal Asiatique* 242. 69-82.

Haudricourt, A.-G. (1954b). Comment Reconstruire le Chinois Achaique. In *Word* 10. 351-64.

Hombert, J.-M. (1978). Consonant Types, Vowel Quality, and Tone. In Victoria Fromkin (ed.), *Tone: A Linguistic Surv*ey. New York: Academic Press. 77-111.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan (1979). Phonetic Explanations for the Development of Tones. In *Lg* 55:1. 37-58.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis & S. Winters (1999). A Cross-linguistic study of stop place perception. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. 2069-2072.

Hume, E. & K. Johnson (2003). The Impact of Partial Phonological Contrast on Speech Perception. In *Proceedings of the XVth International Congress of Phonetic Sciences*.

Hume, E. & M. Seo (2004). From Speech Perception to Optimality Theory: Metathesis in Faroese and Lithuanian. In *Nordic Journal of Linguistics* 27:1. 35-60.

Hura, S. L., B. Lindblom, & R. L. Diehl (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59-72.

Jakobson, R., Fant, G., and M. Halle (1952) *Preliminaries to speech analysis*: the distinctive features and their correlates. Cambridge: Acoustics Laboratory, Massachusetts Institute of Technology.

Jakobson, R. & M. Halle (1956). *Fundamentals of language*. Gravenhage: Mouton & Co.

Janson, T (1983). Sound change in perception and in production. In *Lg* 59:1. 18 – 34.

Johnson, K. (1988). Processes of Speaker Normalization in Vowel Perception. Unpublished PhD dissertation, Department of Linguistics, The Ohio State University.

Johnson, K. (2004). Cross-linguistic perceptual differences emerge from the lexicon. In Agwuele, A., Warren, W. and S.-H. Park (eds.) *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception.* Sommerville, MA: Cascadilla Press. 26-41.

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

Kohler K. (1990). Segmental reduction in connected speech: Phonological facts and phonetic explanations. In Hardcastle, W.J. & A. Marchal (eds.) *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Publishers. 69-92.

Kratochvil, P. (1968). *The Chinese language today: features of an emerging standard*. London: Hutchinson.

Krishnan, A., Xu, Y., Gandour, J. & P. Carian (2005). Encoding of pitch in the human brainstem us sensitive to language. In *Cognitive Brain Research* 25. 161-168.

Kuhl, Patricia K. (1991). Human adults and human infants show a "perceptual magnet effect" for the protypes of speech categories, monkeys do not. In *Perception & Psychophysics* 50:2. 93-107.

Lee, Y-S, Vakoch, D.A., & L. H. Wurm (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research* 25. 527-542.

Liljencrants, J. & B. Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. In *Lg* 48:2. 839-862.

Macmillan, N. A. (1987). In Harnad, S. (ed.) *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Maran, L. (1973). On becoming a tone languages: a Tibeto-Burman modal of tonogenesis. In L. Hyman (ed.), *Consonant Types and Tone* (Southern California Occasional Papers in Linguistics, No.1). 97-114.

Maspero, H. (1912). Etudes sur la phonetique historique de la langue annamite: les initiales. In *Bulletin de l'Ecole Française d'Extreme Orient* 12.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (ed.), *Consonant Types and Tone* (Southern California Occasional Papers in Linguistics, No.1). 71-95.

Mielke, J. (2003). The Interplay of Speech Perception and Phonology: Experimental Evidence from Turkish. In *Phonetica* 60:3. 208-229.

Miyawaki, K., W. Strange, R. Verbrugge, A. M. Lieberman, J. J. Jenkens & O. Fujimura (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. In *Perception and Psychophysics* 18. 331-340.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrik, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior: CLS*. Chicago: Chicago Linguistics Society. 178-203.

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

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.

Schlaug, G., L. Jaencke, Y. Huang & H. Steinmetz (1995). In vivo evidence of structural brain asymmetry in musicians. In *Science* 267. 699-701.

Seo, M. (2001). A Perception-based Study of Sonorant Assimilation in Korean. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (The Ohio State University working papers in linguistics, No. 55). 43-69.

Stagray, J. R., & D. Downs, 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. In *Journal of Chinese Linguistics* 21:1. 143-163.

Steriade, D. (2001). A perceptual account of directional asymmetries in assimilation and cluster reduction. In E. Hume & K. Johnson (eds.) *The role of perception in phonology*. New York: Academic Press.

Stiles, J. (2000). Neural Plasticity and Cognitive Development. In *Developmental Neuropsychology* 18:2. 237-272.

Svantesson, J.-O. (2001). Tonogenesis in Southeast Asia – Mon-Khmer and beyond. In Shigeki Kaji (ed.), *Proceedings of the Symposium: Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Japanese Accentology, and Other Topics*. ILCAA, Tokyo University of Foreign Studies. 45-58.

Trubetzkoy, N. S. (1969). *Principles of phonology*. C. Baltaxe (translator). Berkeley and Los Angeles: University of California Press. (Translated from N. S. Trubetzkoy, 1939. *Grundzuger der Phonologie*. Travaux du Cercle Linguistique de Prague 7, Prague.)

Tserdanelis, G. (2001). A Perceptual Account of Manner Dissimilation in Greek. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (The Ohio State University working papers in linguistics, No. 55. 172-199.

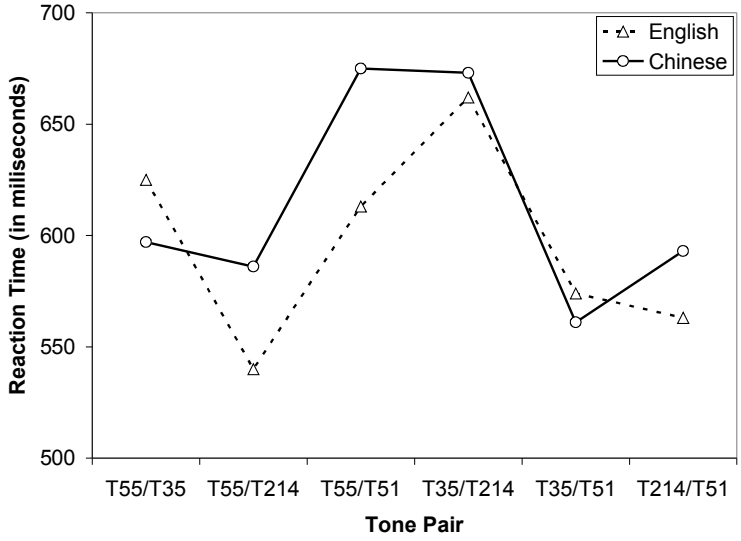
Wang, S-Y. W. (1976). Language change. In *Annals of New York Academy of Sciences*. 61-72.

Werker, J. F. and R. C. Tees (1984). Phonemic and phonetic factors in adult cross-language speech perception. In *JASA* 75:6. 1866-1878.

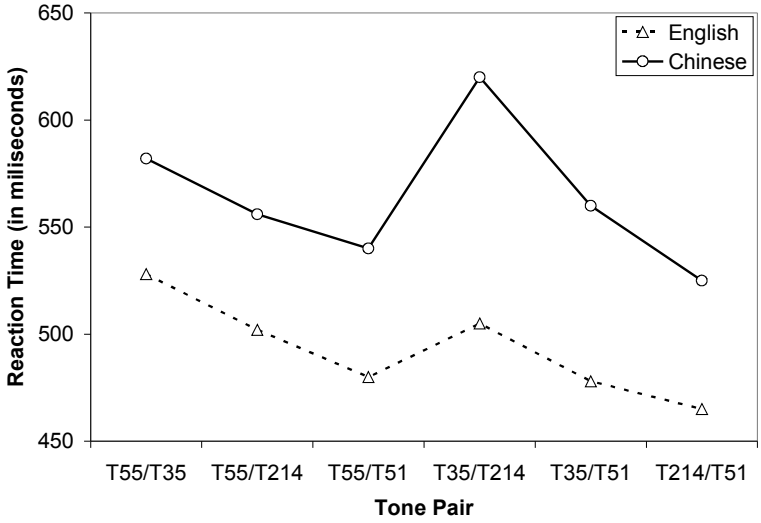
Werker, J. F. and J. S. Logan (1985). Cross-language evidence for three factors in speech perception. In *Perception & Psychophysics* 37:1. 35-44.

Appendices

A: Collapsed plot of RT data from Experiment 1, ignoring order in a pair of tones (Cf. Figure 2). The oddity in the synthetic in T51 may have sounded unnatural to Chinese listeners.



B: Collapsed plot of RT data from Experiment 2, ignoring order in a pair of tones (Cf. Figure 3). There is a notably large and significant between-subject RT difference for pairs T35/T214.



C: Collapsed plot of difference rating data from Experiment 3, ignoring order in a pair of tones (Cf. Figure 4). On this five-point scale, smaller numbers are assigned to more similar tones, while larger numbers indicate dissimilarity. It is obvious from this plot that for AE listeners, only the high vs. low contrast in T55/T214 and T51/T214 was salient, while for Chinese listeners, only the tones involved in a sandhi rule (i.e. T35/T55 and T35/T214) were similar.